AD-A253 071

Tech. Rep. STI-TR-8817-006
June 1992

DTIC
ELECTE
JUL 2 0 1992
S C D

BIG GRAPHICS AND LITTLE SCREENS: MODEL-BASED DESIGN OF LARGE SCALE INFORMATION DISPLAYS

Paul R. Frey
William B. Rouse

Search Technology, Inc.

Atlanta, GA

Rosemary D. Garris

Naval Training Systems Center
Orlando, FL

Prepared for
Office of the Chief of Naval Research
810 North Quincy Street
Arlington, Virginia 22217-5000

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

June 1992

92-19064

92 7

This report was prepared under the Navy Manpower, Personnel, and Training R&D Program of the Office of the Chief of Naval Research under Contract N00014-89-C-0047.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
June 1992

3. REPORT TYPE AND DATES COVERED
Final (1 Jan 89 - 30 Jun 92)

4. TITLE AND SUBTITLE
Big Graphics and Little Screens: Model-Based Design of Large Scale Information Systems

5. FUNDING NUMBERS
N00014-89-C-0047

6. AUTHOR(S)
Paul R. Frey
William B. Rouse
Rosemary D. Garris

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Search Technology
4725 Peachtree Corners Circle, Suite 200
Norcross, GA 30092

8. PERFORMING ORGANIZATION REPORT NUMBER
STI-TR-8817-006

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
STI-TR-8817-006

11. SUPPLEMENTARY NOTES
Supported by the Office of the Chief of Naval Research Manpower, Personnel, and Training R&D Program

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The design of large scale information displays is addressed. Problems with traditional approaches to display design are discussed. It is argued that the evolving nature of humans' roles in complex systems will exacerbate these problems. A model-based framework for display design is proposed involving system models, task models, and humans' models of systems and tasks. This framework provides a basis for exploring three types of display design problems, including problems of evolution, deviation, and change. Use of the overall conceptual framework is illustrated in the context of an application involving design of computer-based graphical displays for a maintenance information system. Traditional maintenance information includes large graphical drawings that are difficult to portray on the small screens of computer-based maintenance information systems. This research investigates the design of graphical displays using display abstraction and aggregation as design parameters. A display's aggregation level reflects the field of view of a display, such as component, assembly, or system level diagrams. A display's abstraction level reflects the representation contained in a diagram, such as a component's form, function, or purpose in an assembly. The results from five experiments with experienced maintenance personnel are presented and design guidelines are suggested.

14. SUBJECT TERMS
displays, troubleshooting, job performance aids, maintenance diagrams, graphical material

15. NUMBER OF PAGES
104

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Tech. Rep. STI-TR-8817-006
June 1992

BIG GRAPHICS AND LITTLE SCREENS: MODEL-BASED DESIGN OF LARGE SCALE INFORMATION DISPLAYS

Technical Report STI-TR-8817-006

Approved by:

Paul R. Frey
Program Manager

William B. Rouse
Principal Investigator

Ruston M. Hunt
President

DTIC QUALITY INSPECTED 8

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

Tech. Rep. STI-TR-8817-006
June 1992

BIG GRAPHICS AND LITTLE SCREENS: MODEL-BASED DESIGN OF LARGE SCALE INFORMATION DISPLAYS

Paul R. Frey
William B. Rouse

Search Technology
Atlanta, GA

Rosemary D. Garris
Naval Training Systems Center

Orlando, FL

ABSTRACT

The design of large scale information displays is addressed. Problems with traditional approaches to display design are discussed. It is argued that the evolving nature of humans' roles in complex systems will exacerbate these problems. A model-based framework for display design is proposed involving system models, task models, and humans' models of systems and tasks. This framework provides a basis for exploring three types of display design problems, including problems of evolution, deviation, and change. Use of the overall conceptual framework is illustrated in the context of an application involving design of computer-based graphical displays for a maintenance information system. Traditional maintenance information includes large graphical drawings that are difficult to portray on the small screens of computer-based maintenance information systems. This research investigates the design of graphical displays using display abstraction and aggregation as design parameters. A display's aggregation level reflects the field of view of a display, such as component, assembly, or system level diagrams. A display's abstraction level reflects the representation contained in a diagram, such as a component's form, function, or purpose in an assembly. The results from five experiments with experienced maintenance personnel are presented and design guidelines are suggested.

This report was prepared under the Navy Manpower, Personnel, and Training R&D Program of the Office of the Chief of Naval Research under Contract N00014-89-C-0047.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

The design of information displays has long been an important problem. Increasingly information-rich environments now make it even more important. Consequently, problems associated with display design, if not resolved, are likely to be more and more troublesome.

One problem is the typical early lack of specificity. During conceptual design, humans' likely tasks are often ill-defined, resulting in information requirements being undefined and the design of effective displays being virtually impossible. As the design of the system evolves, the details of tasks and information requirements emerge and display design becomes possible. However, it often is too late in the design process to modify tasks, and hence requirements, if the display design process uncovers potential performance problems.

A second problem is the changing nature of tasks. One aspect of this concerns the ways in which task definitions evolve as design progresses. Of more fundamental importance is the general changes that humans' tasks have undergone in recent years. Humans' roles have shifted from manual control to monitoring, decision making, and problem solving. As a result, much of task performance no longer involves overt activity. This makes it difficult to determine information requirements.

A third problem concerns the lack of a principled approach to designing information displays. Those current approaches that could legitimately be termed principled require levels of design detail that preclude solution of the above problems. What is needed is a principled approach that is useful during both conceptual and detailed design.

This goal is best elaborated by discussing it in the context of earlier approaches to display design. The time-honored, task analytic approach relies on

complete specification of information and control activities for which requirements are derived. Based on these requirements, display elements are chosen and integrated into display pages or panels.

Considerable effort has been invested in formalizing this process by using the human factors research literature to derive display principles and guidelines, e.g., Smith & Mosier (1986). Structured display design procedures have been developed that explicitly incorporate these principles and guidelines, either manually (e.g., Frey, Sides, Hunt & Rouse, 1984) or via computer-based support (e.g., Frey & Wiederholt, 1986; Hunt & Frey, 1987).

In general, these approaches have not been as successful as had been anticipated. One reason is that the set of codifiable principles and guidelines is not as comprehensive as had been imagined. Available rules emphasize perception of information via single display elements, and say little about integrated displays and pictorial presentations. Another problem with available principles and guidelines is the limited extent to which context-free rules can sufficiently specify context-specific displays. Beyond these problems, even the rules that are available often involve attributes that cannot be assessed by a computer and, consequently, limit the applicability of computer-based support.

In order to move beyond these problems, a new approach is needed. This approach should enable consideration of the task context within which information displays will be used. Further, to the extent possible, the way in which task context is considered should be representable in a form that can be accessed and manipulated by a computer-based support system.

This report discusses progress on development of a model-based framework for achieving these objectives. This framework differentiates among three types of models. One type involves one or more representations of the system -- equipment, people, and the organization. A second type involves one or more representations

of tasks -- goals, plans, and associated actions. The third type of model is the human's models of the system and tasks.

The framework outlined in the next section of this report emphasizes the interactions of these three types of models. This enables delineation, in a subsequent section, of three classes of design problems, as well as appropriate methods for dealing with these problems. The modeling concepts introduced and design problems illustrated in these general discussions are subsequently illustrated in the context of an application involving design of displays for a maintenance information system.

II. MODEL-BASED FRAMEWORK

Figure 1 provides a high-level depiction of the components of the model-based framework. The system model is a representation of the system in terms of block diagrams, signal and data flow graphs, mockups, simulators, etc. The task model is a representation of tasks via production or performance goals, procedures, operational sequence diagrams, mode control logics, etc.

The system and task models represent the way in which the equipment, people, and organization were designed, configured, supported, and intended to operate. The human's models are the ways in which a human represents, explicitly or implicitly, the system and tasks. These models underlie the human's ability to perform tasks acceptably within the system context (Rouse & Morris, 1986).

The three types of models in Figure 1 provide a basis for determining information requirements. If it can be assumed that the human's models are equivalent to the system and task models, then information requirements analysis can be based almost solely on knowledge of the system and task designs -- humans can be assumed to conform to these constraints.

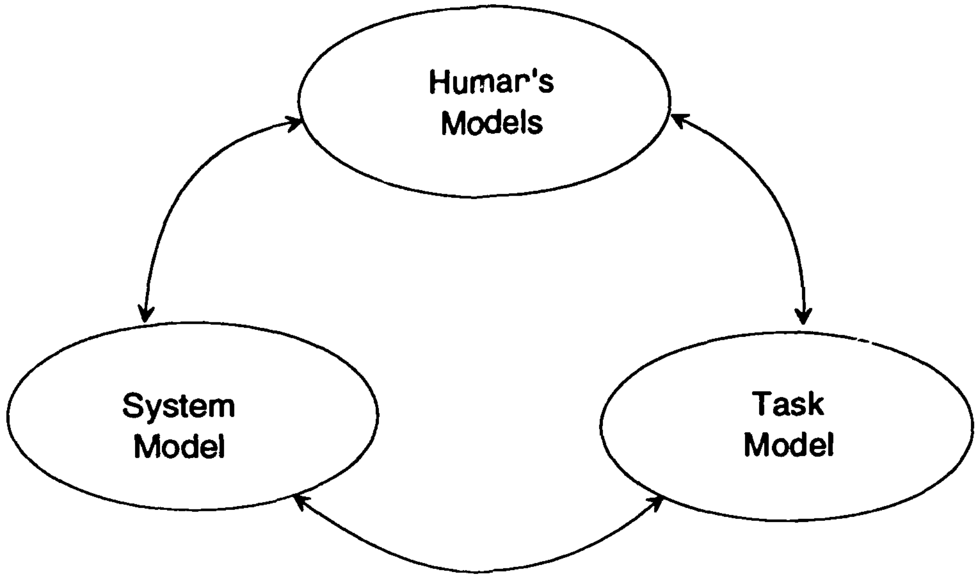

Figure 1. Component Models of Framework

While this assumption is quite commonly adopted, if only tacitly, it is obvious that increased system complexity renders the assumption untenable. Consequently, the human's models have to be viewed as other than identical to system and task models. This leads to the possibility of many alternatives. The range of alternatives can be considered in terms of the extent of system knowledge and task knowledge necessary for acceptable performance.

Figure 2 depicts the possible range of system knowledge. The elements of this figure concern how the system works. Knowledge ranges from the simple identity of system components, to how elements co-function, to the principles and theories that underlie the functioning of the system. At one time, it was thought that personnel needed a good knowledge of all the elements of Figure 2 in order to be competent operators, maintainers, etc. However, numerous experiments by a wide range of investigators have shown this intuition to be incorrect -- see Morris & Rouse (1985a, 1985b) and Rouse & Morris (1986) for summaries of these studies. In

general, knowledge of elements toward the upper left of Figure 2 are most important, and elements toward the lower right are less important.

LEVEL	TYPES OF KNOWLEDGE		
	"WHAT"	"HOW"	"WHY"
Detailed/ Specific/ Concrete	Characteristics of System Elements (What Element Is)	Functioning of System Elements (How Element Works)	Requirements Fulfilled (Why Element Is Needed)
↓	Relationships Among System Elements (What Connects To What)	Co-Functioning Of System Elements (How Elements Work Together)	Objectives Supported (Why System Is Needed)
Global/ General/ Abstract	Temporal Patterns Of System Response (What Typically Happens)	Overall Mechanism Of System Response (How Response Is Generated)	Physical Principles/Theories (Why: Physics, Chemistry, Etc.)

Figure 2. System Knowledge

Figure 3 illustrates the possible range of task knowledge. The elements of this figure concern <u>how to work the system</u>. Knowledge ranges from knowing what can happen, to how to deal with situations, to the principles and theories upon which procedures and strategies are based. High levels of abilities in applying task knowledge are termed skills.

LEVEL	TYPES OF KNOWLEDGE		
	"WHAT"	"HOW"	"WHY"
Detailed/ Specific/ Concrete	Situations (What Might Happen)	Procedures (How to Deal With Specific Situations)	Operational Basis (Why Procedure is Acceptable)
↓	Criteria (What is Important)	Strategies (How To Deal With General Situations)	Logical Basis (Why Strategy is Consistent)
Global/ General/ Abstract	Analogies (What Similarities Exist)	Methodologies (How To Synthesize And Evaluate Alternatives)	Mathematical Principles/Theories (Why: Statistics, Logic, Etc.)

Figure 3. Task Knowledge

As with system knowledge, not all of the elements of Figure 3 are required for successful performance. The elements toward the upper left tend to be more important than those toward the lower right. For both system and task knowledge, elements in the lower right are most important when personnel are expected to deal with unforeseen or novel situations.

A mental model is a popular construct for characterizing humans' system knowledge and, to an extent, task knowledge. The nature of mental models is depicted in Figure 4. Succinctly, mental models are the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states (Rouse & Morris, 1986).

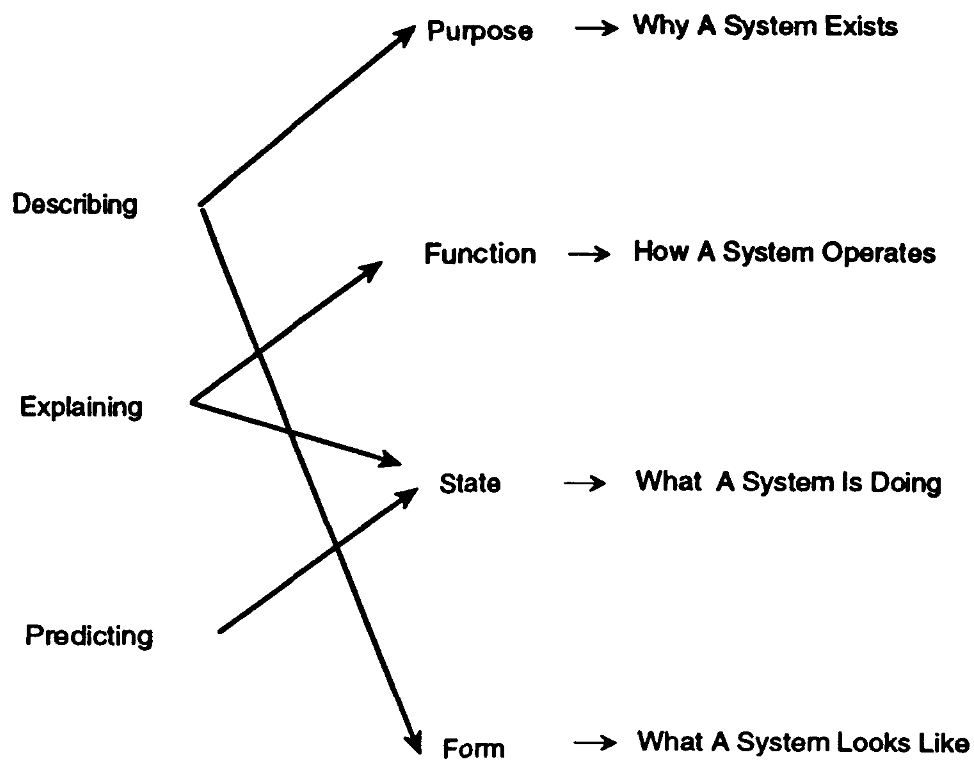

Figure 4. Nature of Mental Models

This construct can serve two important purposes. First, information requirements can be based on supporting a human's mental model. For example, the ways in which a human explains and predicts system state have implications for information requirements concerning values of variables, current modes, status of failures, etc. Put simply, information displays can be designed to be compatible with humans' mental models.

The second important use of the mental models construct relates to the need to foster appropriate mental models. Rather than attempting to support inappropriate mental models, such models should be remediated via aiding, training, or some combination. This possibility is elaborated in the next section.

The notion of the mental models is intuitively appealing. However, there are many open issues surrounding this concept. A central issue concerns alternative representations of purpose, function, and form (e.g., equations, rules, scripts, frames, and images). Another issue concerns the extent to which displays should support humans' mental models vs. facilitate development of particular mental models. In other words, to what extent should displays be adapted to humans vs. displays fostering adaptation by humans? This represents another perspective on the aiding vs. training issue noted above.

III. TYPES OF DESIGN PROBLEMS

Conceptualizing information requirements analysis and display design in terms of the three types of models in Figure 1 enables defining three general classes of design problem. By defining these problems in terms of models (M), it is possible to delineate alternative approaches to resolving these problems.

A. Problems of Evolution ($M^* \rightarrow M$)

As depicted in Figure 5, conceptual design is typified by evolving system and task models. The system model evolves (i.e., $S^* \rightarrow S$) which, in turn, results in an evolving task model (i.e., $T^* \rightarrow T$). Consequently, information requirements continue to evolve (i.e., $R^* \rightarrow R$). Thus, the "what" of display design is a moving target.

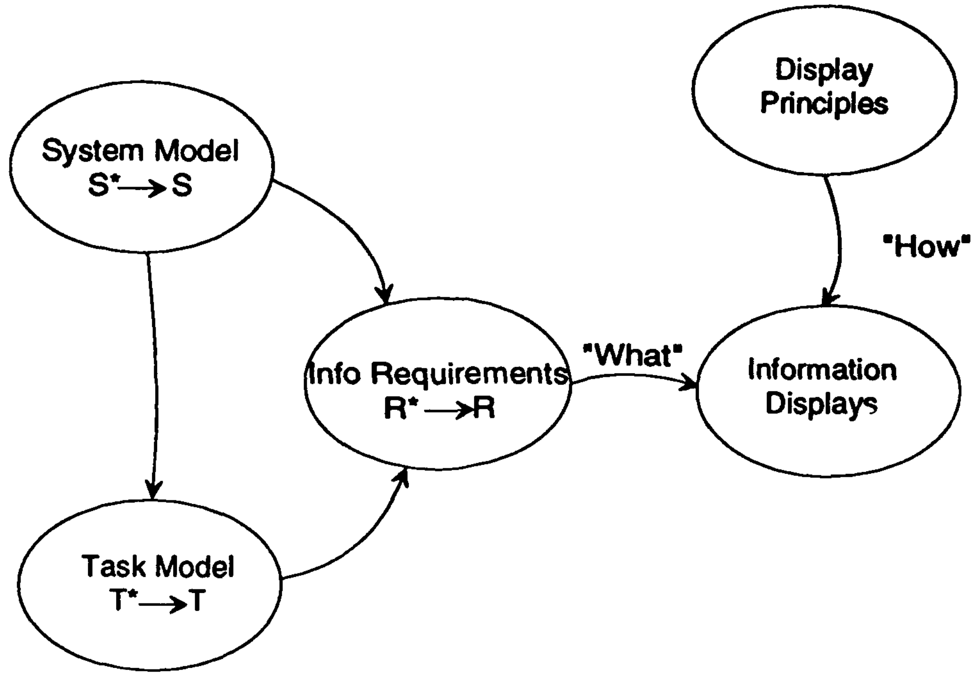

Figure 5. Evolutionary Nature of Models

Within the context of this evolutionary process, it is difficult to consider the human's models. Even if one makes the simplifying assumption that the human's models are identical to S and T, it is very difficult to consider the "how" of display design when "what" has yet to be settled. The natural tendency is to wait until the evolutionary process has tapered off and then deal with the information displays. As noted earlier, this can be too late.

There are several approaches for dealing with problems of evolution early in the design process. By developing structured design documents and databases, the design process can be monitored and audited. If the evolving information requirements are crisply linked to the objectives and functions underlying the system and task models, it is relatively painless to regularly update displays based on model changes, or modify models due to reactions of users to displays (Hunt & Frey, 1987; Rouse, et al., 1990; Rouse & Hammer, 1990).

The latter mode of evolution (i.e., where requirements drive models) can be greatly facilitated by using prototypes to obtain comments and suggestions from likely display users. This process can result in users helping, perhaps indirectly, in the evolution of S and T. It is important to note that without the aforementioned structured design databases, it is often difficult to determine why displays look as they do. Consequently, it can be extremely difficult to determine how S and T should change to yield R, and hence displays, consistent with users' comments and suggestions.

In general, the evolutionary process will be smoother, and model convergence more efficient, if the design process involves all the stakeholders in the process and the product (Rouse, 1991). Thus, beyond users, the design process should also involve customers (or purchasers), technical reviewers, and other people who intentionally influence the function and form of the eventual product. If this process is supported by appropriately structured documentation and databases, problems can be identified, debated, and resolved expeditiously.

B. Problems of Deviation ($\hat{M}$ - M)

Figure 6 depicts four ways in which a human's model ($\hat{M}$) might deviate from the system and/or task model (M). As noted earlier, incompleteness and a degree of inaccuracy are inevitable for complex systems and tasks. Incompatibilities imply a mismatch between a human's models and the information requirements dictated by the system and task models. Incorrectness is seldom a problem with experienced personnel, but can be troublesome with entry-level personnel who, for example, tend to act on the basis of "naive physics."

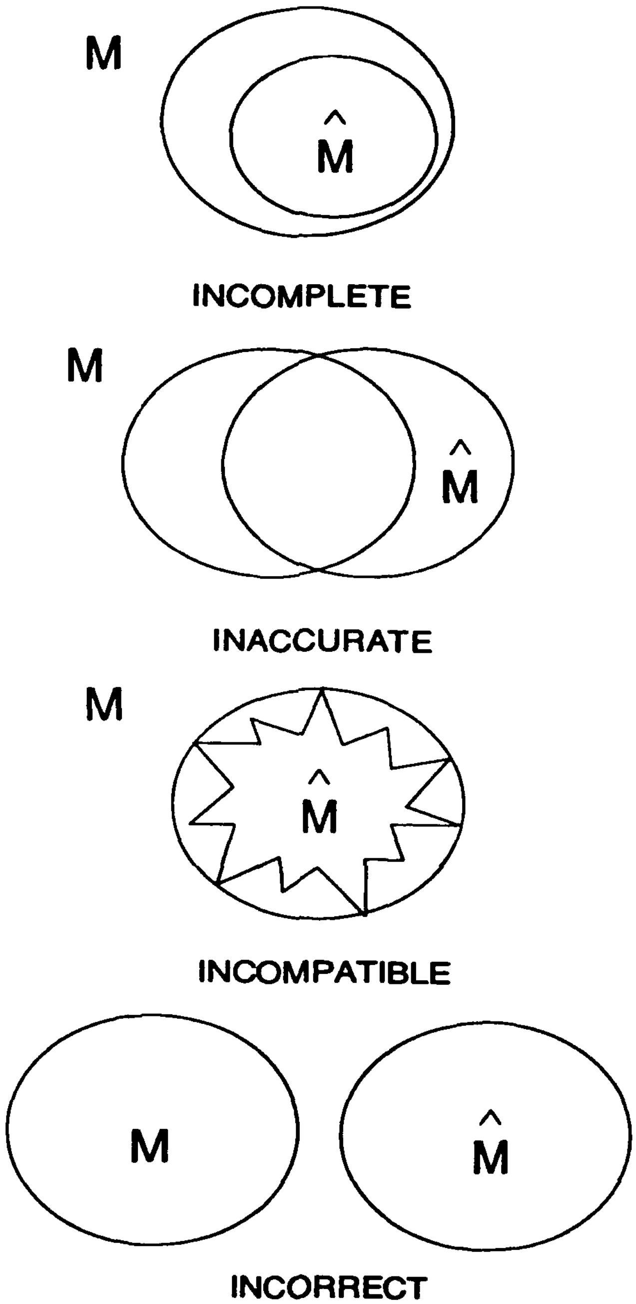

Figure 6. Types of Model Deviation

The impacts of model deviations on display design are summarized in Figure 7. If humans must conform to S and T, then R can be determined based solely on S and T. Variations in $\hat{T}$, but not $\hat{S}$, imply that humans understand the system, but are

free to reconceptualize their tasks. As a consequence, displays based on T can be incompatible with successful performance via $\hat{T}$. Variations in $\hat{S}$, but not $\hat{T}$, are not very troublesome with stable designs. However, changes of S may not influence $\hat{S}$ and hence $\hat{T}$ in predictable ways and task performance will be fragile relative to design changes. Finally, if both $\hat{S}$ and $\hat{T}$ can vary freely, display design is an extremely difficult endeavor.

Task Model / System Model		Allowable Variation of Human's Model ($\hat{T}$-T)	
		Low	High
Allowable Variation of Human's Model ($\hat{S}$-S)	Low	Easy to Determine R Based Solely On S and T	Easy to Determine "What;" More Difficult To Determine "How"
	High	Easy to Determine R; More Difficult to Vary R Based on S	Difficult to Determine R Based Solely On S and T

Figure 7. Impact of Model Deviations

Several approaches are useful for dealing with problems of deviation. The use of procedures can force $\hat{T}$ to conform to T and, to a much lesser extent, $\hat{S}$ to S. Training is also very important, particularly if deviations can be expressed in terms of

deviations of one or more of the elements of Figures 2 and 3. A particularly attractive alternative is embedding training and/or aiding in an operational system to remediate model deficiencies when they are encountered. Thus, rather than dealing with all possible deficiencies in advance, one only deals with actual deficiencies as they occur (Rouse, 1987).

C. Problems of Change (ΔM) - Big Graphics and Little Screens

A third class of problems concerns changes of S and T due to system updates, task modifications, or use of new technologies. While there are a wide variety of ways in which this can happen, a very common change within the realm of display design is the transition from hardcopy displays to computer-generated displays, as well as transitions from large-screen stationary displays to small-screen portable displays. These types of transition present interesting problems, and this will be the primary focus of the remainder of this report. Contrary to intuition, simply transitioning information from, for example, hardcopy technical manuals to computer displays is not necessarily an improvement (Rouse & Rouse, 1980; Morehead & Rouse, 1983).

1. Prior Research on Paging, Scrolling, and Windowing Large Graphical Displays

Compared to the parallel presentation of graphics in traditional paper-based drawings, multipage displays inherently present information serially. This has been shown to result in more errors, especially in more complex systems (Geiser & Schumacher, 1976). Error rates can potentially be reduced by display design techniques such as grouping and integration. A study by Mitchell and Miller (1983) evaluated a grouping scheme that did not improve error rates and an integration scheme that was successful. The integration scheme not only grouped data based

on user's tasks, but also presented pre-processed information that was more compatible with the user's high level information needs.

Paging through multipage displays is not the only way a large amount of information can be arranged and viewed. One can organize the information as a single, very large page. To view all of the information, one can *scroll* the page under the display screen. Alternatively, one can *window* the display screen over the page. A comparison of paging, scrolling, and windowing found that paging leads to fewer errors than scrolling (Schwarz, Beldie & Pastoor, 1983), and windowing is faster and leads to fewer errors than scrolling (Bury, et al., 1982). Duchnicky and Kolers (1983) report data on the readability of scrolled text as a function of line length, character density, and window height.

It is common for multiple display pages to be arranged hierarchically. An important issue for such displays concerns the number of levels in the hierarchy, which tends to be traded off against the amount of information per display page (that is, less information per page leads to more pages, which often leads to more levels in the hierarchy). The amount of information per page is constrained by the display size. Two studies by Henneman and Rouse (1984a & 1984b) compared two- and three- level hierarchies in terms of fault diagnosis performance in large scale dynamic networks. They found substantial degradations in performance for the larger number of levels. Duncan (1982) compared hierarchical paging to scrolled displays for a static fault diagnosis task. He found that the hierarchical-paged display is somewhat better than the scrolled display and that both displays lead to less variance in performance than with a single display showing the entire system. A considerable amount of effort is being devoted to studying such "hypermedia" systems (Glushko, 1989).

Another approach to viewing large graphical displays on small computer screens is to eliminate unnecessary detail. This technique has been used in the

aerospace industry to simplify the design and use of paper-based technical documentation (Aerospace Industries Association, 1989). Results of research examining level of detail in computer-based graphical displays have indicated that detail levels lower than 100% can be as effective as full detail in supporting maintenance task performance and learning. In addition, evidence suggested that different types of graphic information may be required depending on whether the task was trained or aided (Garris, Mulligan, Ricci, Dwyer, McCallum & and Moskal, 1990a; Garris, Mulligan, Ricci, & McCallum, 1990b; Ricci, Garris, & McCallum, 1990). "Fisheye views" can be used to dynamically change the level of detail in a given region of a graphic displayed on a computer screen (Furnas, 1986).

Considering the design of individual display pages, a wealth of human factors data is available. Reviews by Smith and Mosier (1986), Frey and colleagues (1984), and Tullis (1983) provide guidelines for design and evaluation of computer-generated displays. Several design tools have been developed based on these guidelines (Flanagan, Blue, Giacaglia, Lenorovitz, & Stanke, 1987; Frey, 1989; Perlman & Moorhead, 1988; Tullis, 1986). Most of the existing guidelines and tools provide guidance on the detailed design of display formats, but they do not provide direction on the initial selection of display formats. It is expected that the current effort will demonstrate that it is necessary to explicitly consider the user's tasks in the selection of display formats, leading to task-specific guidelines for selecting formats for displays.

2. Abstraction and Aggregation

Figure 8 summarizes several of the ways in which one can display on a small-screen computer information that was previously presented on large hardcopy media. The primary alternatives are scrolling, zooming, and branching. One can zoom displays by increasing or decreasing the field of view (FOV). Displays can be

scrolled by keeping the size of the field of view constant and changing the portion of the system that is displayed. Branching among screens, windows, and elements can involve changes in the field of view; but it can also involve changing the representation (REP) of the system. The currently very popular hypermedia technology provides a means to create display systems that enable branching among windows and elements.

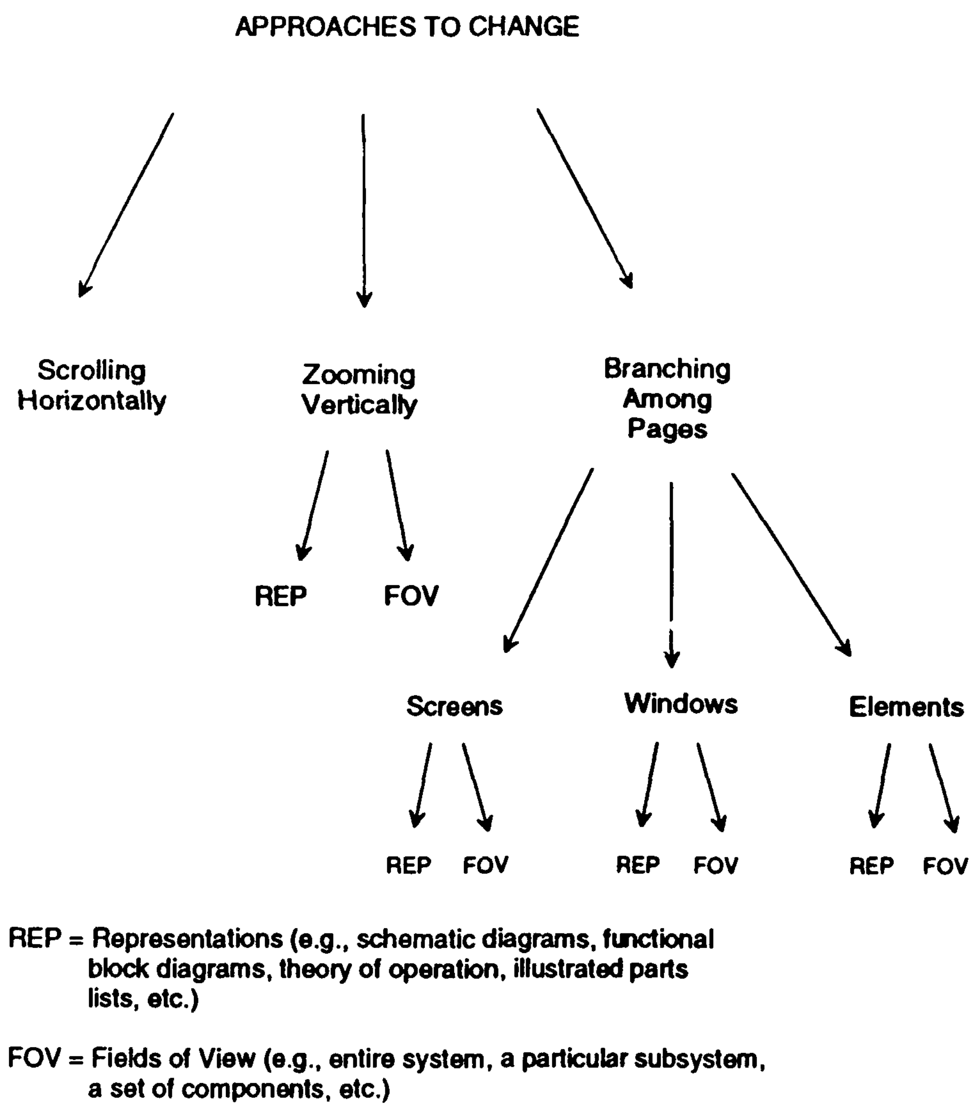

Figure 8. Approaches to Change

Providing different representations can be viewed as an attempt to support the human's mental models in general, and the currently appropriate mental model in particular. Changes in the field of view can be thought of as narrowing or broadening the scope or coverage of the display to match current information needs. Thus, humans can be viewed as invoking different mental models at different points in time, which implies time-varying information requirements. Many different displays can share a single point on the dimensions of representation and field of view, such as displays showing different portions of the system (i.e. different fields of view) at the same level of representation. Field of view relates to level of detail, in that both of these concepts can be used to describe the amount of graphic information contained in a display.

Branching to change representation or field of view for engineering systems can be conceptualized as moving in Rasmussen's (1986) aggregation-abstraction space shown in Figure 9. Changes in the field of view may involve moving along the aggregation dimension (i.e., increasing or decreasing the field of view). Changes in the field of view may also occur without moving along this domain, for example presenting a different part of the system without increasing nor decreasing the field of view. Changes in representation involve moving along the abstraction dimension.

Experience has shown that Figure 9 is very useful for categorizing displays. Further, displays at different points in the aggregation-abstraction space may produce differences in human performance, which is illustrated in the results of Experiment One in the next section. Thus, the levels of aggregation and abstraction of displays do matter. This conclusion does not, however, indicate the most appropriate levels of aggregation and abstraction. The most appropriate levels of aggregation and abstraction are likely to depend upon several factors such as the type of task being performed, the nature of the system being displayed, the experience level of the user, and the nature of the display technology. Some of

these factors are investigated in Experiments Two through Five, which are also described in the following section.

Aggregation					
Abstraction	System	Subsystem	Assembly	Sub-assembly	Component
Functional purpose Production flow models System objectives Constraints					
Abstract function Causal structure, mass, energy, and information flow topology					
Generalized function "Standard" functions and processes: feedback loops, heat transfer					
Physical function Electrical, mechanical, chemical processes of components and equipment					
Physical form Physical appearance and anatomy; material and form; locations					

Figure 9. Aggregation vs. Abstraction Space
(From Rasmussen, 1986)

IV. EXAMPLE APPLICATION

As engineered systems become more numerous and complex, the amount of documentation necessary to support the maintenance of these systems increases dramatically. Furthermore, technology-driven and cost-driven solutions to maintenance are resulting in fewer people performing maintenance on a wider range of equipment. As people are asked to maintain systems that are less familiar to them, there is a heavier reliance placed upon the documentation of those systems. Thus, not only is the amount of documentation increasing, but its importance is increasing as well.

The growth of various technologies holds promise for providing tools and methods for managing the growth of documentation. High density storage devices enable portable systems to contain and access volumes of information while requiring only a few cubic inches of space and weighing only a few pounds. Advances in micro-electronics and software technology enable flexible, interactive, high-speed access to this information. New display technologies are providing light-weight, low-power, high-resolution capabilities to these systems as well. Even with the promise of these capabilities, problems stand in the way of broad application of these technologies to maintenance documentation.

The design issue addressed in this research is the problem of producing computer-based visual displays of maintenance information traditionally portrayed in large-scale, paper-based graphical drawings. This can be succinctly described as the "Big Graphics - Little Screen" (or BGLS) problem.

Display size and resolution are two major parameters that contribute to this problem. Technology limitations currently constrain the resolution of display devices to be 10 to 100 times less than the printed page. But even when the limitations of

display resolution cease to be a major constraint, portability requirements will continue to constrain the size of the display.

The goal of this application was to develop and evaluate principles to guide the transformation of hardcopy formats of technical information to electronically displayed formats. The primary motivation for this transfer was the need to go from large blueprint sized hardcopy, often called C size, to small computer display sized images.

The domain of application for the experiments discussed here was maintenance of the blade fold system of the Navy's SH-3 helicopter. The blade fold system on the SH-3 enables the main rotor blades to be folded and stowed after landing to conserve deck or hanger space. The conservation of space on aircraft carriers is an important goal, the blade fold system can cause significant problems if it fails to function properly. The blade fold system is necessarily complex due to the severe consequences of inadvertent operation. An extensive network of electrical and hydraulic interlocks is required to ensure safe operation of the aircraft. Furthermore, many components of the blade fold system are located on the rotary wing head and are subjected to severe rotational strain as well as vibration and corrosive sea spray. This combination of critical functionality, complex design, and harsh operational environment results in a system that is sometimes trouble-prone and frequently difficult to maintain.

The subjects for these five experiments were SH-3 Aviation Electricians (AE) from HS WING ONE at the Jacksonville Naval Air Station (NAS JAX). Due to the limited available population of trained SH-3 AE's, the number of subjects in each experiment was limited; therefore, care must be taken in generalizing the results. However, the rich context of the experiments (a complex blade fold system, an extensive maintenance information system, trained maintainers, and 3 to 4 hour

exposure times) gives credibility to the application of our results in real environments.

A. Experiment One (Pilot Study)

The primary objective of this initial study was to determine if displays that varied in terms of aggregation and abstraction would differentially affect SH-3 maintainers' performance. Maintainers were given three troubleshooting problems to solve. The first problem was solved using the Navy's standard hardcopy maintenance materials for the SH-3. These materials include several large (11 x 17 inch) sheets of paper showing location and schematic diagrams of the components of the blade fold system. The second and third problems were solved using smaller (8.5 x 11 inch) hardcopy displays that were designed based on the concepts in Figure 9. These displays include three abstraction levels; physical form, physical function, and general function. The diagrams of physical functions include block diagrams and schematics aggregated by the function of the different circuits of the blade fold system. The diagrams of general functions include block diagrams showing interactions among subsystem functions. The participants were six SH-3 maintenance technicians. Three were relatively inexperienced, and the remaining three were much more experienced.

The number of displays that each subject examined during each problem was normalized to a baseline for each problem. The baseline was drawn from standard troubleshooting guides for each of the problems. An analysis of variance showed that the experimental displays developed using the concepts of aggregation and abstraction resulted in a significant reduction in the number of displays examined by maintainers relative to conventional maintenance information (see Figure 10, $F(2,12) = 18.491$, $p < .001$). There was also a significant interaction effect between

experience and display material ($F(2,12) = 5.334$, $p < .05$). Experienced maintainers examined more displays relative to the baseline than inexperienced maintainers when using standard materials. However, when using the experimental displays, experienced maintainers examined fewer displays relative to the baseline than did the inexperienced maintainers.

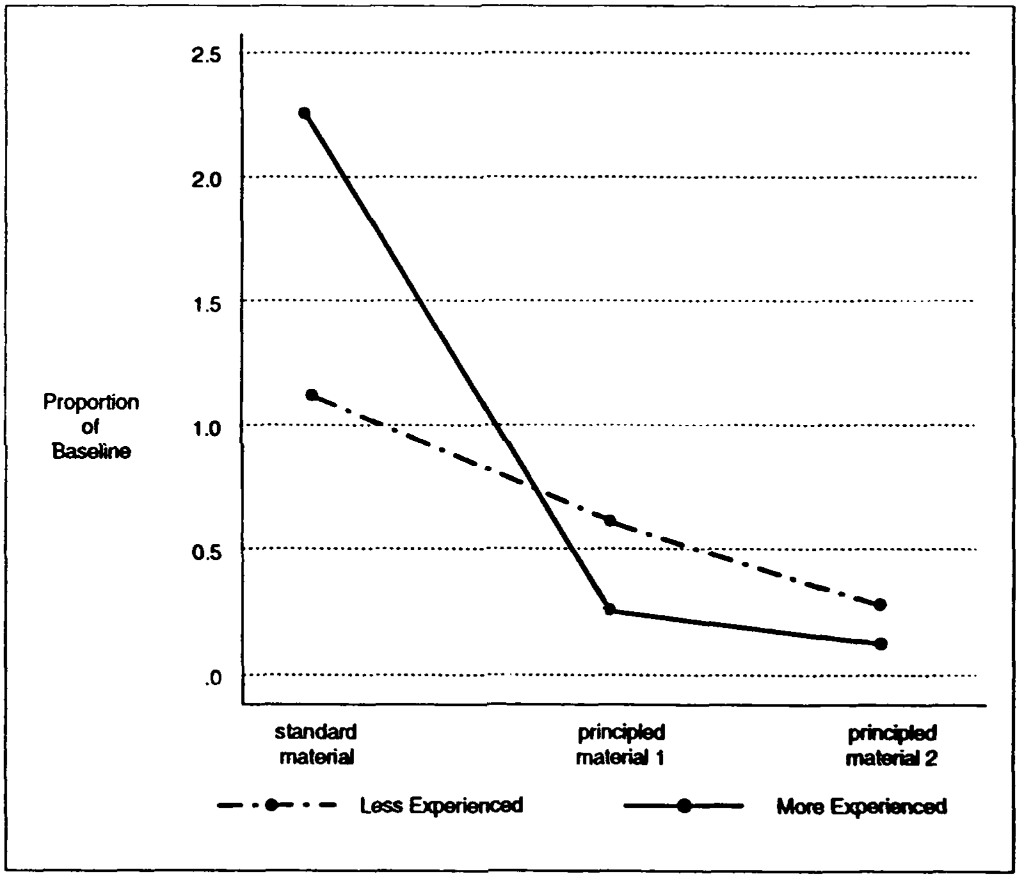

Figure 10. Number of Displays Examined Relative to Baseline (Experiment 1)

These results showed that the experimental displays have the potential to enhance maintenance performance. While it is difficult to draw strong conclusions from this initial study, it showed that human performance may be differentially

affected by levels of abstraction and aggregation. For a more detailed description of the dependent variables and the results, see Sewell, Rouse & Johnson, 1989.

B. Experiment Two

The primary objective of Experiment Two was to determine if the maintainers' use of display aggregation/abstraction levels varied with the type of maintenance task or with the experience level of the maintainer. Other major objectives of Experiment Two were to use computer-based displays in the evaluation rather than hardcopy displays and to use a somewhat larger number of subjects.

At a very coarse level, tasks can be simply described as requiring different relative emphasis on "thinking" and "doing." While these categories are difficult to quantify, they can be used to describe qualitative differences between typical maintenance activities addressed in these studies. Figure 11 shows some hypothesized design principles using this as an initial task dimension. These principles specify likely relationships among displays. To illustrate, many maintenance troubleshooting tasks are likely to involve a sequence of thinking and doing activities. Initial hypothesis formulation is likely to involve more thinking than doing, while the testing of hypotheses typically requires more doing than thinking. The principles in Figure 11 suggest that such tasks would best be supported in the following way:

- o First viewing abstract, aggregate displays (e.g., functional block diagrams at the system level),
- o Then moving to less aggregation (e.g., block diagrams at the subsystem and assembly levels),

- o Then moving to less abstraction (e.g., schematics at the subsystem and assembly levels),

- o Then moving to less aggregation (e.g., schematics at the sub-assembly and component levels), and

- o Finally moving to low abstraction and aggregation (e.g., physical form at the component level).

ATTRIBUTES / TASK	CHANGE OF REPRESENTATION	CHANGE OF FIELD OF VIEW
THINKING o Inferring o Deducing o Interpreting o Deciding	Likely to Benefit from More Abstraction	Likely to Involve Movement to Less Aggregation
DOING o Navigating o Locating o Observing o Manipulating	Likely to Require Much Less Abstraction	Likely to Involve Movement Among Levels of Aggregation

Figure 11. Initial Design Principles

1. Maintenance Tasks

To investigate the hypotheses shown in Figure 11, maintenance tasks were selected that require different relative amounts of "thinking" and "doing." Obviously these dimensions are very difficult to quantify in a practical setting, and categorizing these tasks in this manner was subjective. Domain relevance was a major consideration in the selection of these tasks.

The three types of maintenance tasks selected for Experiment 2 were "following procedures," "circuit tracing," and "problem solving."

The first task type, "*following procedures*," consisted of three troubleshooting problems that the subjects performed using a fully-proceduralized job performance aid (FPJPA). The FPJPA was a chart prescribing the tests to be performed and the decisions to be made, leading to subsequent tests or conclusions. This task was chosen to represent tasks that require very little "thinking" in the sense of reasoning about the symptom, system, possible tests, or test results. However, the task does require "doing" in the sense of locating test points, prescribing tests, and binary branching (deciding whether a condition exists or not) in the FPJPA based on the results of the test.

In the "following procedures" tasks, the FPJPAs were given to subjects on paper, and the subjects were told to follow the procedure in solving the problem. In these problems, they located and identified the test points (which were specified in the FPJPA) on the location diagrams. In addition, they were allowed to use any intermediate displays they desired to reach the location diagram showing the physical form and location of the test points. For example, the subjects would frequently find a test point on a schematic diagram, and use the links among displays to access the needed location diagram. Once the test point had been located, the experimenter reported the results of the test, and the subject would proceed to the next step in the procedure.

On the other end of the "thinking/doing" dimension are the "*problem solving*" tasks. These three problems were designed to require the subjects to reason about the system's design, function, and operation. The "problem solving" tasks consisted of determining the operational symptom given the failure of a certain device, identifying a half-split test point given a failure symptom, and troubleshooting a failure without using a FPJPA. In the troubleshooting problem, the subjects performed tests in the same manner as in the "following procedures" tasks, except that they had to decide which tests to perform and interpret the results of the test.

The third task type was "*circuit tracing*." These tasks were selected to fall between the other two task types in the "thinking/doing" dimension. In these problems, the subjects were told to trace various portions of electrical circuits as if they were testing to isolate an open circuit, and locate the test points using the location diagrams. These problems required the subjects to think about the configuration of the circuit, but not the functional operation of the system.

2. Maintenance Information System

For the second experiment, a maintenance information display system was developed for the blade fold system for the SH-3 helicopter. Figure 12 shows the partitioning of the abstraction/aggregation display space for the second experiment. The abstraction space was divided into three levels: flow, schematic, and location diagrams. The aggregation space was also divided into three levels: low (e.g., components or subassemblies), medium (e.g., assemblies), and high (e.g., systems). Forty-five displays were implemented.

The lowest level of abstraction is the location diagrams. These diagrams illustrate the physical form and location of assemblies and components of the blade fold system. An example is shown in Figure 13.

Level of Abstraction \ Level of Aggregation	High	Medium	Low
High (Flows)	1 diagram listing flow diagrams grouped by function	11 diagrams of electrical and hydraulic flow	(no diagrams in this cell)
Medium (Schematics)	1 diagram listing schematic diagrams grouped by function	14 diagrams of function-oriented schematics	(no diagrams in this cell)
Low (Locations)	1 diagram showing entire helicopter and location of major assemblies	3 diagrams showing major assemblies and locations of subassemblies	14 diagrams showing subassemblies and locations of components

Figure 12. Abstraction/Aggregation Space (Experiment 2)

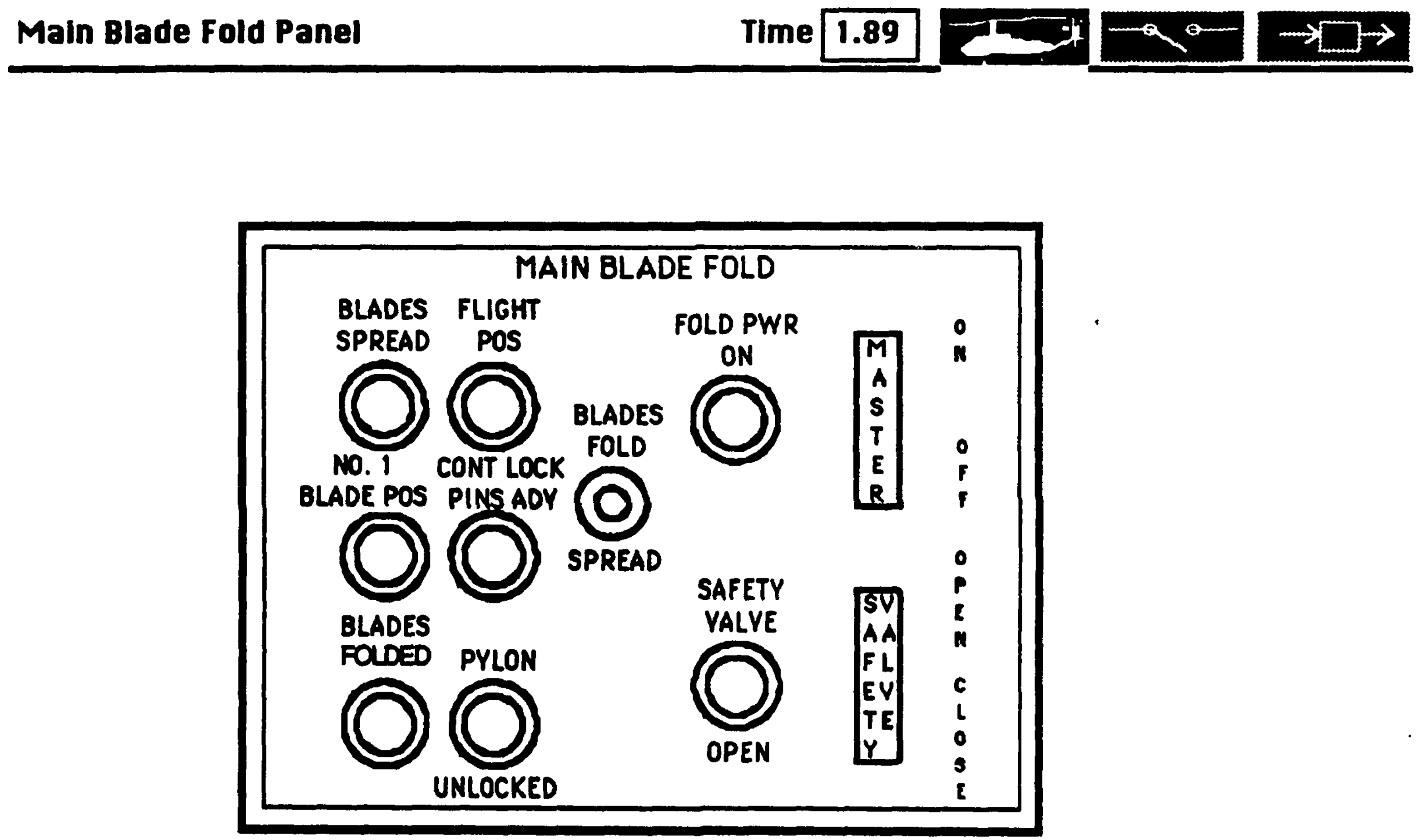

Figure 13. Example of Low Abstraction/Low Ag,regation Display

The next highest level of abstraction is the schematic diagrams. These diagrams are similar to paper-based schematics used in the training courses for maintenance of the blade fold system. An example is shown in Figure 14. They are unlike the diagrams used by the maintenance personnel in the shop, which are essentially large engineering drawings. Partitioning these large drawings into pieces that can be displayed on a computer screen is difficult and is an example of the kind of problem addressed in this report.

The highest level of abstraction in the displays is the flow diagrams. These are block diagrams developed to provide information about the electrical or hydraulic flow in the system without much of the detail contained in the schematic diagrams. An example is shown in Figure 15.

The flow and schematic diagrams were partitioned according to circuit function, resulting in diagrams that contained all of the devices and all of the connections necessary to depict the function of a circuit; however, each device was not fully rendered in any given diagram. For example, a schematic diagram may show only the portion of an electrical relay that is related to the function depicted.

The design of the maintenance information system includes a human/computer interface that allowed the subjects to choose displays in a variety of ways. From any given display, subjects could move to related displays at any level of abstraction or aggregation. Subjects could also choose from a list of displays recently viewed. These options provided the subjects with a great deal of flexibility in selecting displays within the aggregation/abstraction display space.

3. Subjects

The subjects for the experiment were active duty Navy personnel (E-3 through E-6), trained and experienced in maintaining the SH-3 helicopter blade fold system. The subjects were based with HS WING ONE at the Naval Air Station in

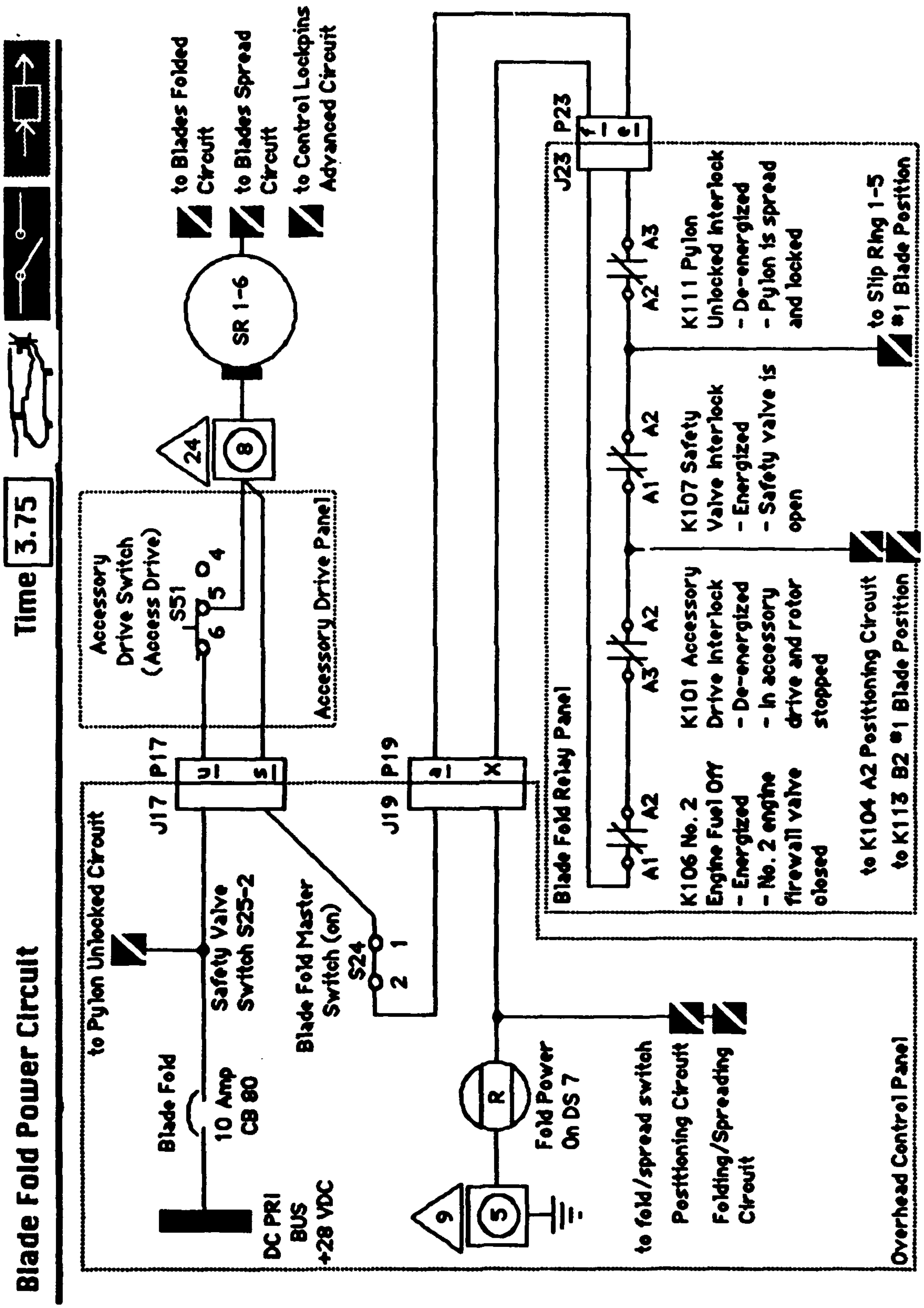

Figure 14. Example of Medium Abstraction/Medium Aggregation Display

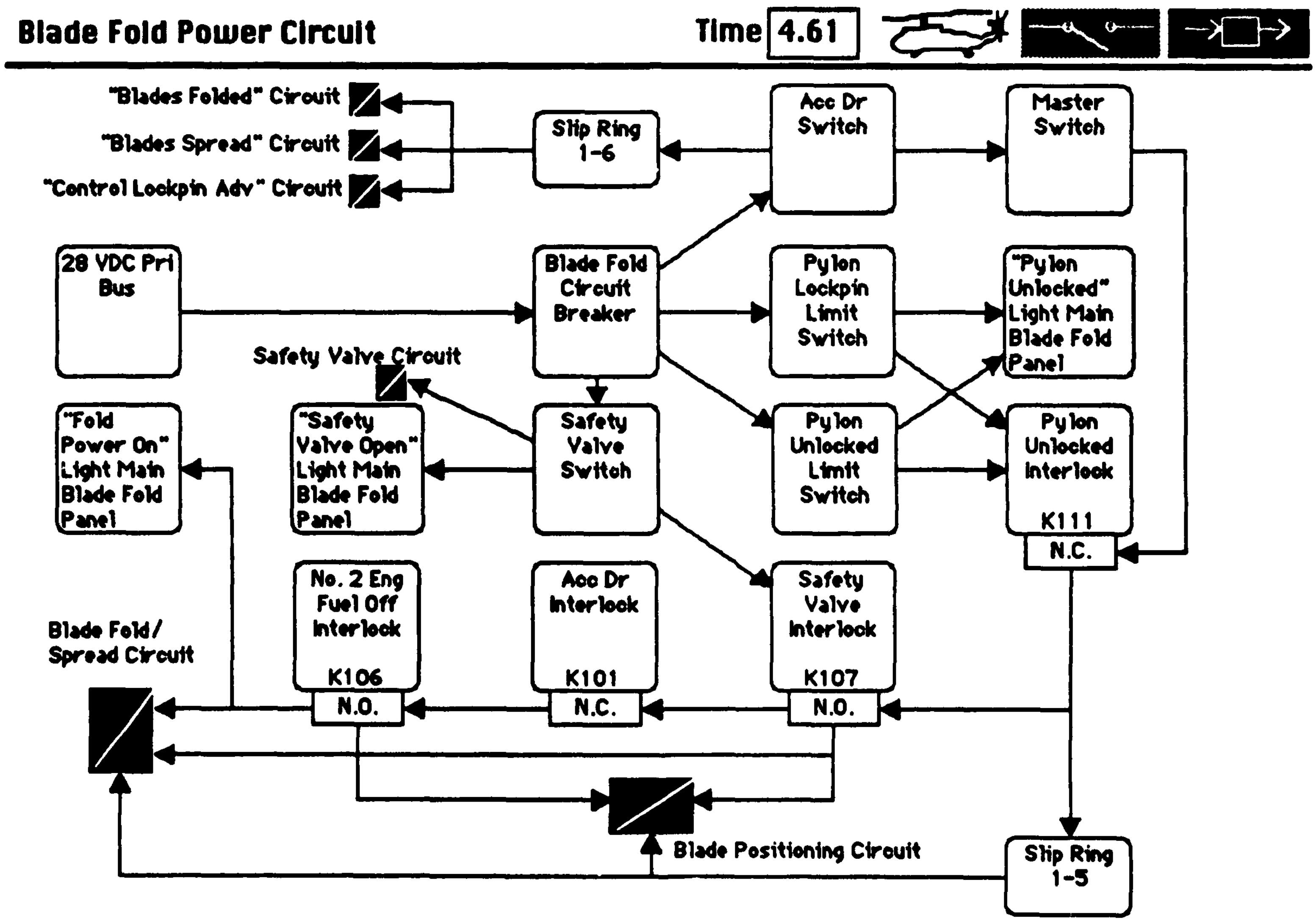

Figure 15. Example of High Abstraction/Medium Aggregation Display

Jacksonville, Florida. The subjects were divided into two groups based on the length of time that they had worked with the SH-3. The less experienced group had less than four years experience, while the more experienced group had four or more years of experience. Each group had five subjects for a total of ten subjects.

4. Measures

Transaction files were collected during each trial. The files contained a list of the displays that the subject accessed, the order in which they were viewed, and the length of time that they were displayed. Based on these transaction files, the following measures were collected: 1) total time on the problem, 2) display usage time for each of the three abstraction levels, 3) display usage time for each of the three aggregation levels, and 4) usage time for displays that were inappropriate for the trial (errors).

At the conclusion of all of the trials, the subjects answered an opinionnaire that addressed several issues related to both the usability of the maintenance information system and their experience with current maintenance documentation.

5. Experimental Design

The experiment used a three factor design with two within-subjects variables and one between-subjects variable. The between-subjects variable was experience level with two groups as defined earlier. The within-subjects variables were problem type and trial number. Each subject received three trials in each problem type for a total of nine trials per subject. The trials were balanced to eliminate any bias due to order of presentation.

Before the trials, each subject received approximately 45 minutes of individual training and practice on the display system. Each subject received the same training and practice. The training included an explanation of the nature of the research, an explanation of each of the display types (abstraction and aggregation levels), and a

structured explanation and demonstration of the operation of the human/computer interface used to select displays. The practice consisted of performing one trial of each task type using symptoms and problems that were not repeated in the experimental trials.

6. Results

The display usage data include the time spent using each display type and errors in display choices. The time data from each trial were standardized as z-scores, means of the z-scores were calculated, and an analysis of variance (ANOVA) was performed to determine the effects of experience level and task type on the usage of the abstraction and aggregation levels of the displays.

Even though the analysis was done using mean standardized scores, in the following description of the results the data are presented as mean percentages of time on the problem. This is done to provide a more meaningful context for discussing the results. Percentages are used rather than times because the three different task types had very different mean times to completion. Therefore the proportion of time spent on different display types is a more meaningful comparison than the actual time. With these caveats in mind, the significant differences described below are generally in the order of magnitude of tens of seconds and in some cases minutes.

Usage of Abstraction Levels. Experience level showed no significant differences in the use of the abstraction levels as a main effect or as an interaction effect with task type. The use of the abstraction levels was significantly affected by task type at all three levels of abstraction (for the use of high abstraction displays, $F(2,16) = 22.3$, $p < .001$; for the use of medium abstraction displays, $F(2,16) = 8.7$, $p < .01$; for the use of low abstraction displays, $F(2,16) = 26.8$, $p < .001$). Since

experience level showed no main or interaction effects, the two experience levels were pooled in the following *post-hoc* analyses of abstraction usage.

On the "problem solving" tasks, the subjects spent 10% of the time using the *high abstraction* displays (the flow diagrams). This was significantly higher than the use of the high abstraction displays on both the "following procedures" tasks (2%, $F(1,8) = 34.6$, $p < .001$) and the "circuit tracing" tasks (2%, $F(1,8) = 18.1$, $p < .005$).

The *medium abstraction* displays (the schematic diagrams) were used 80% of the time during the "problem solving" tasks and 66% of the time on the "circuit tracing" tasks ($F(1,8) = 5.6$, $p < .05$). These were significantly higher than the use of the medium abstraction displays on the "following procedures" tasks (52%, $F(1,8)$, = 6.0, $p < .05$).

The *low abstraction* displays (the location diagrams) were used the most during the "following procedures" tasks (46%). This was significantly higher than the use during both the "circuit tracing" tasks (32%, $F(1,8) = 6.6$, $p < .05$) and the "problem solving" tasks (10%, $F(1,8) = 58.9$, $p < .001$). The difference between the use of the low abstraction displays in the "circuit tracing" and "problem solving" tasks was also significant ($F(1,8) = 21.7$, $p < .005$).

Usage of Aggregation Levels. Experience level showed no significant differences in the use of the aggregation levels as a main effect or as an interaction effect with task type. The use of the aggregation levels was significantly affected by task type at all three levels of aggregation (for the use of high aggregation displays, $F(2,16) = 32.6$, $p < .001$; for the use of medium aggregation displays, $F(2,16) = 7.6$, $p < .01$; for the use of low aggregation displays, $F(2,16) = 25.2$, $p < .001$). Since experience level showed no main or interaction effects, the two experience levels were pooled in the following *post-hoc* analyses of aggregation usage.

The *high aggregation* displays were used the most during the "circuit tracing" tasks (28%). This was significantly higher than the use during both the "following procedures" tasks (15%, $F(1,8) = 53.8$, $p < .001$) and the "problem solving" tasks (10%, $F(1,8) = 50.8$, $p < .001$). The difference between the use of the high aggregation displays in the "following procedures" and "problem solving" tasks was also significant ($F(1,8) = 6.3$, $p < .05$).

The *medium aggregation* displays were used the most during the "problem solving" tasks (86%), which was significantly higher than on the "circuit tracing" tasks (61%, $F(1,8) = 15.6$, $p < .005$) or the "following procedures" tasks (50%, $F(1,8) = 12.0$, $p < .01$).

The *low aggregation* displays were used most during the "following procedures" tasks (35%). This was significantly higher than the use during both the "circuit tracing" tasks (11%, $F(1,8) = 28.3$, $p < .005$) and the "problem solving" tasks (4%, $F(1,8) = 20.7$, $p < .005$).

To more directly test the hypotheses (see Figure 11), three separate analyses were performed to examine the use of the aggregation levels in the first half of the problem (by time) versus the second half. This allows investigation of how the use of aggregation levels changes with time on the problem.

For the "following procedures" tasks, the high aggregation displays were used significantly more in the first half of the problem than in the second half ($F(1,9) = 40.8$, $p < 0.001$), and the low aggregation displays were used significantly less in the first half than in the second half $F(1,9) = 26.9$, $p < .01$). This indicates that the subjects' use of the aggregation levels tended toward the lower aggregation levels later in the problem, which is counter to our hypothesis for tasks requiring more "doing" than "thinking." Possible reasons for this are discussed later.

In the "problem solving" tasks the subjects used the high aggregation displays more in the first half of the problem than in the last half ($F(1,9) = 46.7$, $p < .001$) and

the medium aggregation displays more in the last half than the first half ($F(1,9) = 29.6$, $p < .001$). This movement from the high to the medium aggregation displays generally agrees with our hypothesis for tasks requiring more "thinking" than "doing."

In the "circuit tracing" tasks the high aggregation displays were used more in the first half than the second half ($F(1,9) = 29.3$, $p < .001$).

Errors. The data were analyzed to determine whether experience level or task type influenced the subjects' use of inappropriate displays. An "inappropriate display" is defined as a display that contains no information relevant to the trial. Across all trials, the subjects spent an average of 10% of the time on each problem (or about fifty seconds) viewing displays that were inappropriate for the problem.

There was no significant difference between experience levels, but there was a difference between problem types ($F(2,16) = 5.7$, $p < .05$). On the "problem solving" tasks, the subjects spent only 5% of the time viewing inappropriate displays, which was significantly less than on the "following procedures" tasks (12.1%) or the "circuit tracing" tasks (14.2%).

The results of Experiment Two are discussed in conjunction with Experiment Three following the next section.

C. Experiment Three

The primary objective of Experiment 3 was to refine and extend the design of the maintenance information system and the experimental design based on the results of Experiment 2.

1. Maintenance Tasks

Based on the results of the second experiment, some of the tasks were modified slightly for the third experiment. These were changed to make the individual problems within the task types less diverse.

The "*following procedures*" tasks remained the same as in Experiment 2 and represented the "doing" end of the "thinking/doing" dimension.

The "*problem solving*" tasks were changed completely to make them more representative of common maintenance activities and to make them more uniform within the task type. In the second experiment, the "problem solving" tasks consisted of three different kinds of problems (see the descriptions above). In the "problem solving" tasks in the third experiment, the subjects were given six different symptoms and were told to generate a list of three failures that would produce each of the symptoms. These tasks required the subjects to reason about the symptoms and the system's design and operation.

The third task type in the third experiment consisted of "*troubleshooting*" tasks, in which the subjects searched for the fault causing a symptom without using a FPJPA. The selection of these tasks resulted from the observation of the subjects on the only troubleshooting problem in the second experiment (then under the "problem solving" task type). In the "thinking/doing" dimension, these tasks represent a combination of elements of the other two task types. The "troubleshooting" tasks require that the subjects think about the symptom and suspected faults (as in the "problem solving" tasks) and prescribe tests and interpret results (as in the "following procedures" tasks). These tasks also required subjects to reason about possible tests to perform to isolate the fault.

2. Maintenance Information System

Based on the results of the second experiment, the maintenance information system was modified slightly to provide a simpler linkage between components at

different abstraction levels. This was accomplished by the addition of schematic diagrams organized according to physical devices, which is shown at the medium abstraction/low aggregation level (see Figure 16). An example of one of these displays is shown in Figure 17.

Level of Abstraction \ Level of Aggregation	High	Medium	Low
High (Flows)	1 diagram showing electrical flow among medium aggregation flow diagrams	7 diagrams of electrical flow	(no diagrams in this cell)
Medium (Schematics)	2 lists of diagrams grouped by function and device schematics	10 diagrams of function-oriented schematics	19 diagrams of device-oriented schematics
Low (Locations)	1 diagram showing entire helicopter and location of major assemblies	4 diagrams showing major assemblies and locations of subassemblies	15 diagrams showing subassemblies and locations of components

Figure 16. Abstraction/Aggregation Space (Experiments 3, 4 and 5)

In the maintenance information system used in Experiment 2, there were complex one-to-many linkages between abstraction levels, particularly between the low abstraction diagrams (locations) and the medium abstraction diagrams (schematics). This was a result of designing the schematics as function-oriented diagrams. For example, a single device may perform different operations in more than one system function. Therefore, the diagram showing the physical form and

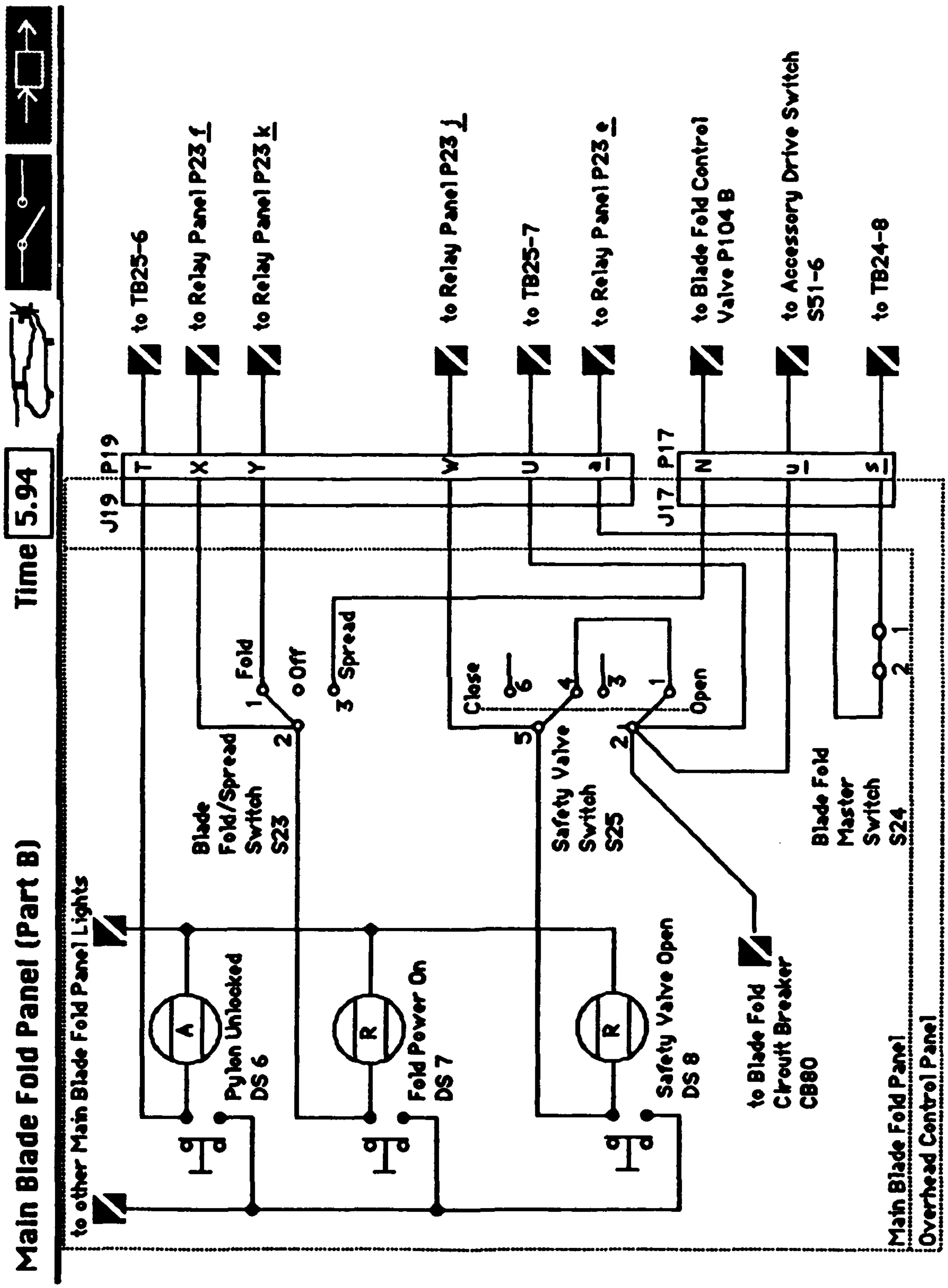

Figure 17. Example of Medium Abstraction/Low Aggregation Display

location of the device (low abstraction/low aggregation) was linked to more than one function-oriented schematic diagram (medium abstraction/medium aggregation). These complex linkages made it difficult for the subjects to find the appropriate schematic diagram based on the location diagrams.

For Experiment 3, the linkages across abstraction levels was simplified by the addition of device-oriented schematics in the medium abstraction/low aggregation cell of the display space. This enabled a simpler one-to-one linkage between the low and medium abstraction levels.

Finally, to improve overall performance of the system, several flow and function-oriented schematic diagrams were removed from the system. These diagrams were not needed for the problems in either experiment, and they were essentially unused in the second experiment.

3. Subjects

As in the second experiment, the subjects for the third experiment were trained, active-duty SH-3 maintenance personnel from HS WING ONE stationed at the Naval Air Station in Jacksonville, Florida. Experiment 3 had six subjects in the more experienced group and seven subjects in the less experienced group for a total of thirteen subjects. Experience level was defined in the same way as in Experiment 2. The less experienced group had less than four years of experience with the SH-3, and the more experienced group had four or more years of experience.

4. Measures

The measures collected in the third experiment were the same as those collected in the second.

5. Experimental Design

The independent variables in Experiment 3 were the same as in Experiment 2 (experience level, problem type, and trial number). In the third experiment, each subject received six trials in each problem type for a total of 18 trials per subject. The training and practice were the same as in the second experiment except for modifications required by the changes in the task definitions. The trials were balanced to eliminate any bias due to order of presentation.

There were two major differences in the way the third experiment was conducted. First, in Experiment 2 the subjects operated the display system, and in Experiment 3 the experimenter operated the display system. This change occurred because of the complexity of the human/computer interface used to select displays. In the second experiment the subjects occasionally "got lost" in the display system and didn't understand how to find their way back to a prior display. Consequently, the data from Experiment 2 included both the displays used in performing the tasks and the displays that the subjects accessed in error. Because this research is addressing display organization and formats rather than display access, it was decided that the experimenter should operate the display system in the third experiment. In this arrangement, the subjects told the experimenter what display they wanted, and the experimenter manipulated the display system to present the display. An on-line method was provided to annotate the subject's data file if a display was erroneously accessed by the experimenter. The total number of erroneous accesses by the experimenter was less than 2% of the nearly 1600 display accesses. The time associated with these accesses was excluded in the subsequent analysis. The subjects' errors in display selection were identified and analyzed off-line.

The second difference in the way the two experiments were conducted was in the required use of the location diagrams. In Experiment 2, the subjects were

required to use the location diagrams to locate each test point once during each trial. After observing the subjects in Experiment 2, it was apparent that they already knew the location of some of the test points on the SH-3 and did not need information about where the test point was or what it looked like. In the third experiment, the experimenter queried the subject about the location of each of the test points. If the subject knew the location already, then the location diagram was not accessed. This should cause the use of the displays in the experimental tasks to more closely resemble their use in the real world.

6. Results

As with Experiment 2, the display usage data (i.e., the time spent using each type of display and time spent using inappropriate displays) from each trial in Experiment 3 were standardized as z-scores, and an ANOVA was performed to determine the effects of experience level and task type on the usage of the abstraction and aggregation levels of the displays. Again, the results are presented as a percentage of time on the problem, and the significant differences reported below are generally in the order of magnitude of tens of seconds and in some cases minutes.

Usage of Abstraction Levels. As in Experiment 2, experience level showed no significant difference as a main effect, and each task type did show a significant main effect on the use of the abstraction levels (for the use of high abstraction displays, $F(2,22) = 29.4$, $p < .001$; for the use of medium abstraction displays, $F(2,22) = 10.0$, $p < .01$; for the use of low abstraction displays, $F(2,22) = 15.0$, $p < .001$).

Unlike Experiment 2, experience level did show an interaction effect with task type (see Figure 18) in the usage of the high abstraction displays (the flow diagrams) ($F(2,22) = 5.0$, $p < .05$). The use of the medium abstraction displays (the schematic

diagrams) reflected a similar interaction of experience level and task type (see Figure 19), although the difference was not statistically significant ($F(2,22) = 2.7$, $p = .086$). In the following description of the results of the *post-hoc* analyses, the two groups are pooled unless explicitly stated otherwise.

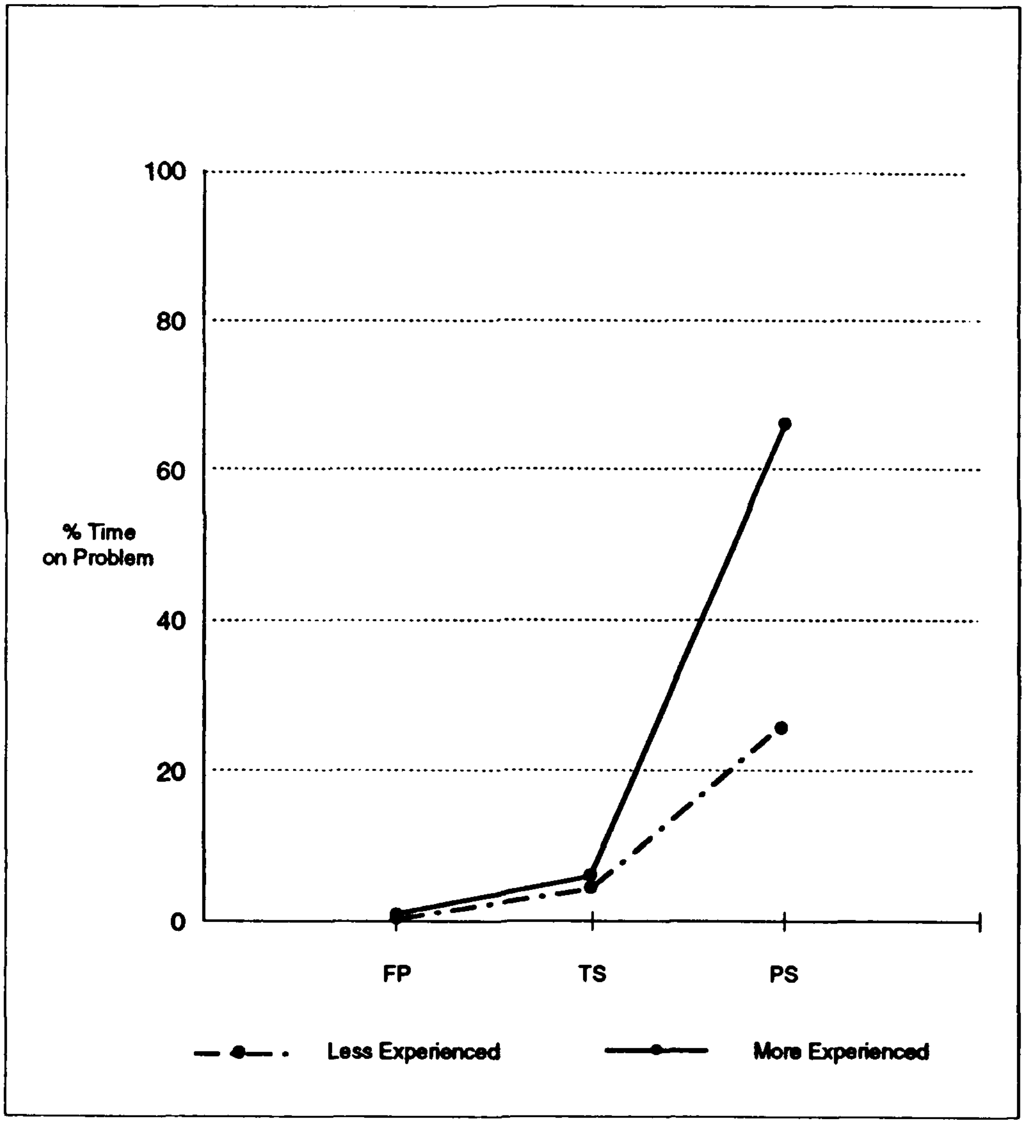

Figure 18. Usage of High Abstraction Displays Affected by Task Type and Experience Level (Experiment 3)

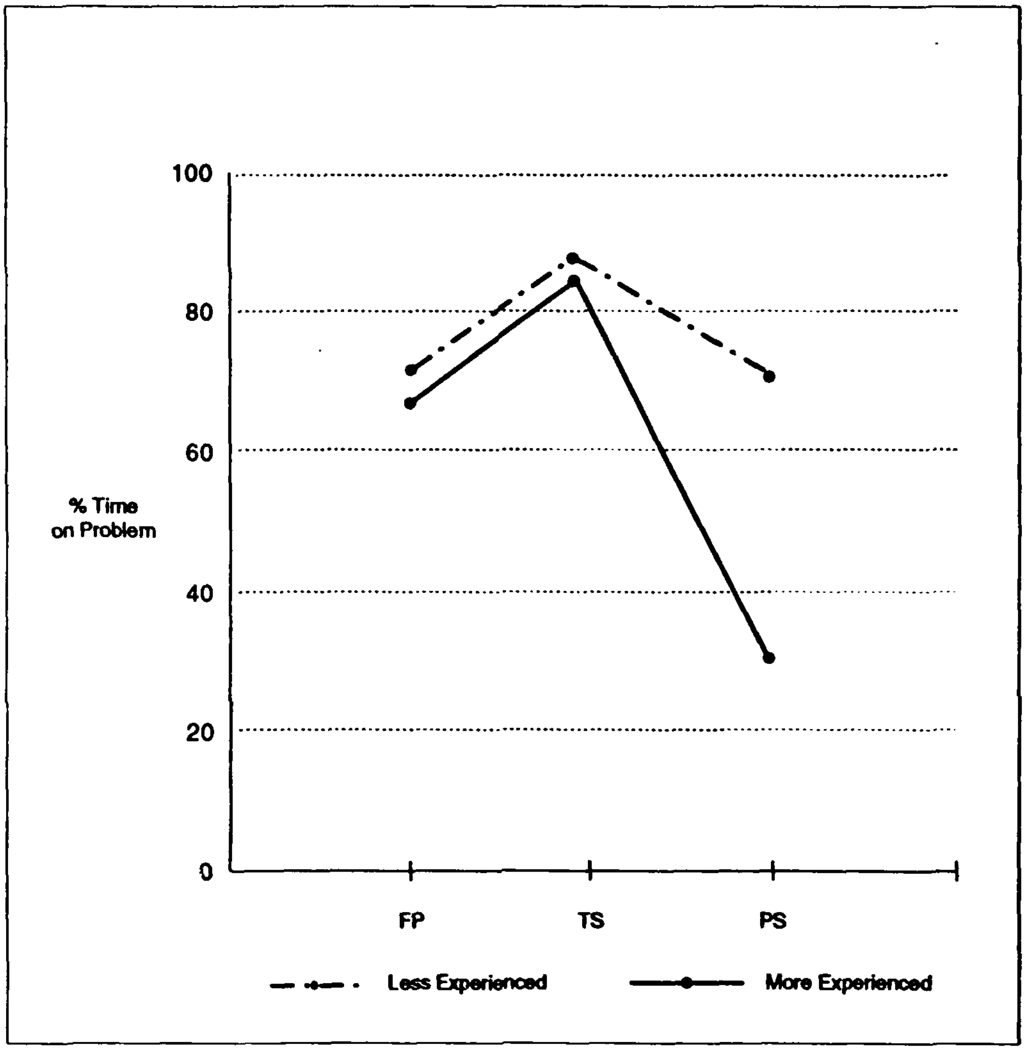

Figure 19. Usage of Medium Abstraction Displays Affected by Task Type and Experience Level (Experiment 3)

As in the second experiment, the *high abstraction* displays were used significantly more on the "problem solving" tasks (43%) than in the other two task types (5%, $F(1,11) = 25.3$, $p < .001$ for the "troubleshooting" tasks; and 1%, $F(1,11) = 34.1$, $p < .001$ for the "following procedures" tasks). However, in Experiment 3

experience level showed an interaction effect with task type. On the "problem solving" tasks, the more experienced group spent 64% of the time on the problem using the high abstraction displays. For the less experienced group, this was 26%, which was significantly lower ($F(1,11) = 4.8$, $p = .05$)

The usage of the *medium abstraction* displays changed somewhat in the third experiment. During the "troubleshooting" tasks, both groups spent approximately 85% of the time using the medium abstraction displays, and during the "following procedures" tasks it was significantly lower, approximately 70% ($F(1,11) = 9.1$, $p < .05$). During the "problem solving" tasks, the more experienced group used the medium abstraction level less (34%) than did the other group (72%), although in this case the difference was not significant. This reflects the difference mentioned in the preceding paragraph.

As in Experiment 2, the *low abstraction* displays (the location diagrams) were used most during the "following procedures" tasks (29%). This was significantly higher than the use of the low abstraction displays on both the "troubleshooting" tasks (10%, $F(1,11) = 15.3$, $p < .005$) and the "problem solving" tasks (2%, $F(1,11) = 18.7$, $p < .005$).

Usage of Aggregation Levels. As in Experiment 2, experience level showed no significant differences in the use of the aggregation levels as a main effect or as an interaction effect with task type. The use of the aggregation levels was significantly affected by the main effect of task type for all three levels of aggregation (for the use of high aggregation displays, $F(2,22) = 9.8$, $p < .01$; for the use of medium aggregation displays, $F(2,22) = 13.9$, $p < .001$; for the use of low aggregation displays, $F(2,22) = 10.9$, $p < .01$). Since experience level showed no main or interaction effects, the two experience levels were pooled in the following *post-hoc* analyses of aggregation usage.

The *high aggregation* displays were used the most during the "following procedures" tasks (26%). This was significantly higher than both the "problem solving" task type (11%, $F(1,11) = 6.8$, $p < .05$) and the "troubleshooting" task type (10%, $F(1,11) = 11.9$, $p < .01$).

The *medium aggregation* displays were used the most during the "problem solving" tasks (85%). This was significantly higher than the "following procedures" tasks (46%, $F(1,11) = 17.2$, $p < .005$) and the "troubleshooting" tasks (70%, $F(1,11) = 6.4$, $p < .05$).

The usage of the *low aggregation* displays was also in agreement with the second experiment. On the "following procedures" tasks, the subjects spent 28% of the time using the low aggregation displays. This was significantly higher than the usage on the "problem solving" tasks (5%, $F(1,11) = 12.7$, $p < .01$). The usage on the "troubleshooting" tasks (20%) was also higher than on the "problem solving" tasks ($F(1,11) = 15.0$, $p < .005$).

As in the analysis of the data from the second experiment, three additional analyses were performed to examine the use of the aggregation levels in the first half of the problem (by time) versus the second half.

For the "following procedures" tasks, the results agreed with the second experiment, which disagrees with the hypothesis. That is, the high aggregation displays were used significantly more in the first half of the problem than in the second half ($F(1,12) = 45.8$, $p < .001$), and the low aggregation displays were used significantly more in the second half than in the first half ($F(1,12) = 62.5$, $p < .001$).

The aggregation use in the "problem solving" tasks also showed the same general pattern as the second experiment, which agreed with the hypothesis. The high aggregation displays were used significantly more in the first half than in the second half ($F(1,12) = 32.3$, $p < .001$). The low aggregation displays were used more in the second half than in the first ($F(1,12) = 19.1$, $p < .01$).

In the "troubleshooting" tasks, the aggregation reflected the same tendency as in the other two task types, which was toward lower aggregation displays with time on the problem. The use of both the high and medium aggregation displays was higher in the first half of the problem than in the second half (for high aggregation, $F(1,12) = 36.1$, $p < .001$; for medium aggregation, ($F(1,12) = 10.0$, $p < .01$). The use of the low aggregation displays was higher in the second half than in the first ($F(1,12) = 25.3$, $p < .001$).

Errors. As with Experiment 2, the data were analyzed to determine whether experience level or task type influenced the subjects' use of inappropriate displays. Across all trials, the subjects spent a much smaller percentage of time (3.5%) viewing inappropriate displays than in the second experiment. Based on the average time on the problems (across task types), this is an average of only eight seconds per trial, which is much less than in Experiment 2 (fifty seconds). This was probably due to a notable difference in the way the trials were conducted. This will be discussed later. For this error measurement, there was no significant difference between experience levels, and only a marginal effect by task type ($F(2,22) = 3.4$, $p = .053$). The differences between task type were so minimally different as to be practically insignificant.

D. Discussion - Experiments 2 and 3

The results generally confirm the original hypothesis about the effects of task type on use of the abstraction levels. The subjects used the high abstraction displays more on the tasks that required thinking about the system's design and operation, and they used the low abstraction displays more on the tasks that involved more of the "doing" aspect of their job.

These results were consistent across the two experiments and across experience levels, although the third experiment showed much greater use of the high abstraction displays and a less frequent use of the medium abstraction displays in the "problem solving" tasks. This difference is most likely due to the differences in the "problem solving" trials between Experiment 2 and Experiment 3. In the second experiment the trials were diverse, whereas in the third experiment they were very uniform. Specifically, in Experiment 2 the "problem solving" trials included a troubleshooting trial which required performing measurements as well as reasoning through the problem. This explanation is supported by noting the similarity between the use of the abstraction levels on the "troubleshooting" tasks in Experiment 3 and the "problem solving" tasks in Experiment 2.

The hypothesis about the effect of task type on the use of the aggregation levels was only partially supported by the results of the experiments. During all tasks, there was a significant tendency to use the high aggregation displays early in the problem and the low aggregation displays later in the problem. The hypothesis was that the tasks which required a greater emphasis on "doing" things would result in movement among the aggregation levels and a uniform use of the different aggregation levels throughout the problem. This did not occur.

One plausible explanation for this tendency in the "following procedures" tasks is that the subjects spent the early stages of a task searching for the initial test points in the first step of the procedure. This searching activity may have required greater use of the high aggregation displays. In the later stages of a task, the subjects may have used the medium and low aggregation schematic diagrams to guide the location of test points, thus requiring less use of the high aggregation displays.

The results from the analysis of errors in display selection were very different in the two experiments. In Experiment 2 the subjects spent a much greater amount

of time (10% per trial, or an average of approximately 50 seconds) on inappropriate displays than they did in the third experiment (3.5% or an average of only eight seconds per trial). The most plausible explanation for this difference is in the way the two experiments were conducted. In Experiment 2, the subjects operated the display system, whereas in Experiment 3 the experimenter operated the system at the direction of the subject. The subjects in Experiment 2 frequently "got lost" and did not have the familiarity with the human/computer interface to quickly recover from incidental errors. In Experiment 3, the subjects could simply direct the experimenter to return to a certain display.

The difference in the errors in display selection probably indicates a deficiency in the design of the display selection portion of our maintenance information system. However, the particular deficiencies of the human/computer interface are of less importance to the current effort than the content of the displays. Certainly, the investigation of display access methods is essential for the development of viable maintenance information systems. However, the purpose of this research was to investigate how to design effective display formats rather than how to provide efficient access to (possibly ineffective) displays.

Many of the display formats provided in the maintenance information system used in this experiment are not currently available to the maintainers of the SH-3. The typical maintenance documentation consists of the location diagrams and the device-oriented schematics (medium abstraction, low aggregation in this design). The flow diagrams and the function-oriented schematics are not available to the maintainers. (However several subjects commented that they frequently sketch equivalent diagrams when troubleshooting the blade fold system.) During both experiments, the subjects used the novel displays more than 70% of the time. Apparently, the subjects preferred the novel displays over those that are most like

their current documents, possibly due to the fact that the novel displays were more suitable for their tasks.

Even though the novel displays were generally preferred, their use did not improve the subjects' performance in selecting appropriate displays. In both experiments, over 80% of the time that the subjects were using inappropriate displays, they were using the novel displays. Although this percentage is slightly higher than the overall use of the novel displays, it does not necessarily indicate that the novel displays are poorer than the traditional. The difference, which amounts to approximately 2% (or 10 seconds) of the average task time in Experiment 2 and less than 1% (or 2 seconds) in Experiment 3, is too small to lead to any conclusions.

E. Experiment Four

Previous experiments indicated that maintainers' tasks influenced their selection of display abstraction and aggregation levels. However, the measures were not designed to indicate whether providing different abstraction and aggregation levels affected maintenance performance. Experiment 4 was designed to study the effects of experience, training, and display abstraction level on simulated maintenance performance.

1. Experimental Design

The experimental tasks for this experiment focused entirely on troubleshooting tasks. In each of eight trials, the subject was given a failure symptom resulting from a single failure, and the subject used the maintenance information system to identify the failure. The experimenter operated the maintenance information system at the direction of the subject and informed the subject of the results of each ot the tests and/or repair actions. Each trial ran until the subject corrected the failure by replacing or repairing the correct part or until 15

minutes elapsed. Trials that ran greater than 15 minutes were omitted from the analysis.

Independent Variables. The experiment used a three factor design with two between-subjects variables and one within-subjects variable.

Experience level. As in the prior experiments, experience level was included as a between-subjects variable. It was expected that maintenance performance measures would show that more experienced subjects perform better than the less experienced subjects.

Training. This between-subjects variable is the training provided to the subjects. Both groups had approximately 45 minutes of individual training and practice on the display system, but the content differed. Both groups were given the same initial training on the types of displays and the organization of the display system. During the second part of the training, one group received a continuation of the first part, emphasizing the form, format, and layout of the displays.

The second group received training emphasizing how to use the different kinds of displays in different situations. The emphasis of this "how to use" training was on the high abstraction displays, which are the most novel displays for these subjects. The training also contained comparisons of the usefulness of the different abstraction levels in different situations. It was expected that this "how to use" training on the displays would result in better maintenance performance, fewer errors, and increased usage of the high abstraction displays.

Availability of high abstraction displays. This within-subjects variable was intended to study whether the high abstraction displays improve maintenance performance. Each subject was given eight trials, four with high abstraction displays and four without. The availability was balanced across trials to reduce bias resulting

from differences in difficulty among the trials. The subjects were informed about the availability of the high abstraction displays at the beginning of each trial. It was expected that experienced subjects and subjects with "how to use" training would perform better on the tasks in which the high abstraction displays were available.

Dependent Measures. The dependent measures were taken from transaction files collected during each trial. These measures included the following:

Maintenance performance (simulated). This is the sum of the actual time the subject spent on the experimental task (time to solution) and a simulated time computed from a library of times for each maintenance operation performed (e.g., time to perform a test and time to repair or replace a part). These times were developed from standard SH-3 maintenance documentation and discussions with domain experts. The library of task times was made available to the subjects, and they were asked to minimize the total time to complete the task. They were provided with a continuous indication of the total simulated task time.

Time to solution. This is the actual time the subjects spent completing a trial.

Diagnostic errors. In each trial, there was only one failed component in the system. An incorrect diagnosis was counted as an error. The subjects received feedback on the correctness of their diagnosis. If their diagnosis was not correct, the subject continued with the trial until they arrived at the correct diagnosis or until the maximum allotted time for each trial (15 minutes) had passed.

Display usage. As in experiments two and three, the amount of time that each subject used each type of display (abstraction and aggregation level) was measured. No feedback was provided to the subject on this measure.

Display errors. Each trial had a pre-defined set of appropriate displays. The amount of time that each subject spent using inappropriate displays was measured, as in the other two experiments. No feedback was provided to the subject on this measure.

2. Subjects

As in the earlier experiments, the subjects were actual maintenance personnel, trained and experienced with the SH-3 blade fold system. The eight subjects in the low experience group were E3's and E4's with 3 to 5 years of maintenance experience. The eight subjects in the high experience group were E5's and E6's with 10 to 15 years of maintenance experience.

3. Results

The data from each trial were standardized as z-scores, and an ANOVA was performed to determine the effects of experience level, training, and abstraction level on the dependent measures. Although these analyses were performed using standardized z-scores, the plots in the following sections are shown in units that are more appropriate for understanding the magnitude of the differences. In most cases, these plots correspond to the plots of the z-scores; however in some cases the plots emphasize differences that are insignificant in the z-scores.

Twenty-three trials (of a total of 128) were excluded from the analysis. Fifteen of these were not completed in the allotted time. Also excluded were eight trials in which the main performance measure (simulated maintenance time) was outside two standard deviations of the mean for the trial. The excluded trials were judged to be outliers and were a mix of experimental conditions with a minimum of nine for the experienced group to a maximum of fourteen for the inexperienced group.

Maintenance Performance. Maintenance performance showed no significant difference in any of the three main effects (experience level, training content, or availability of the high abstraction displays). However, there was a significant interaction between experience level and training content (see Figure 20; $F(1,12) = 8.78$, $p < .05$). The "how to use" training improved the performance of the inexperienced subjects by 8%, which corresponds to roughly 6 minutes ($F(1,6) = 9.02$, $p < .05$). Although Figure 20 suggests a difference in the performance of the experienced subjects based on training, this difference was not significant.

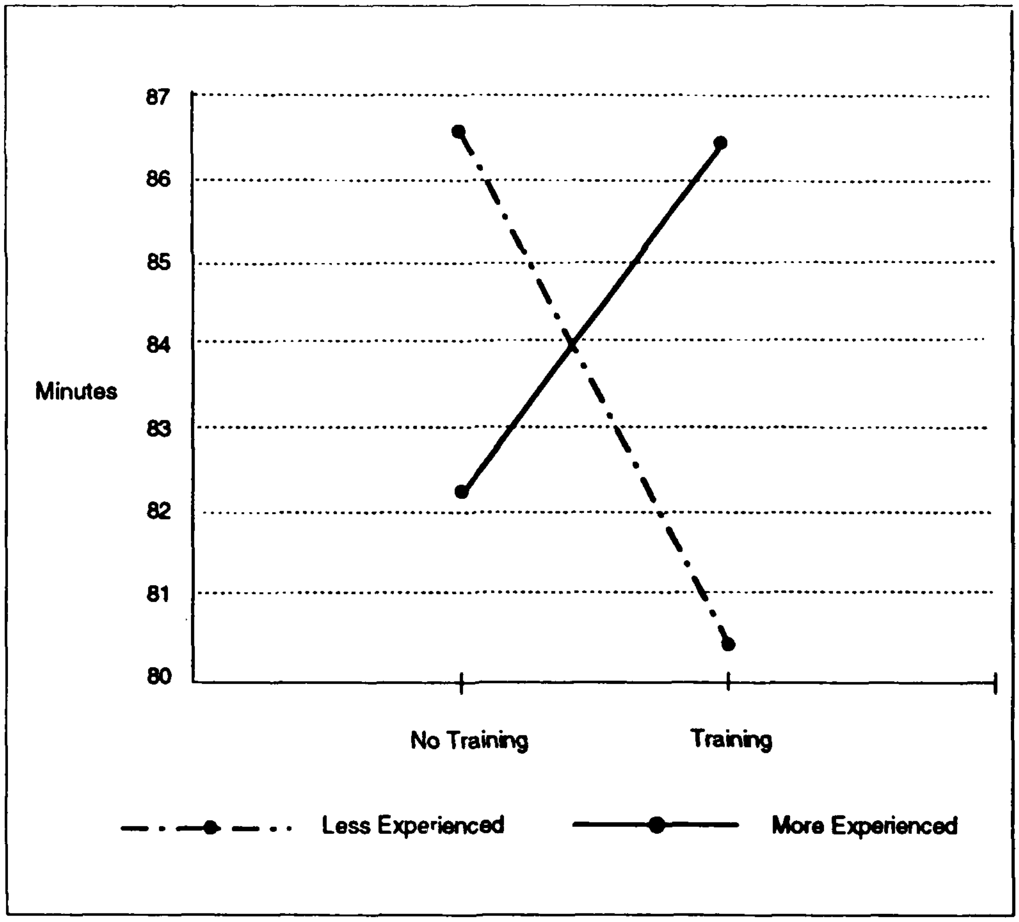

Figure 20. Effect of Experiment and "How to Use" Training on Maintenance Performance (Experiment 4)

There was also an apparent interaction between experience level and availability of the high abstraction displays (see Figure 21), although the result was not statistically significant ($F(1,12) = 4.53$, $p = .055$). Unexpectedly, the inexperienced subjects performed better on trials without the high abstraction displays, although again the result was not statistically significant ($F(1,6) = 5.75$, $p = .053$). The performance of the experienced subjects was not affected by the availability of the high abstraction displays, and the differences in training did not influence performance in trials with the high abstraction displays available.

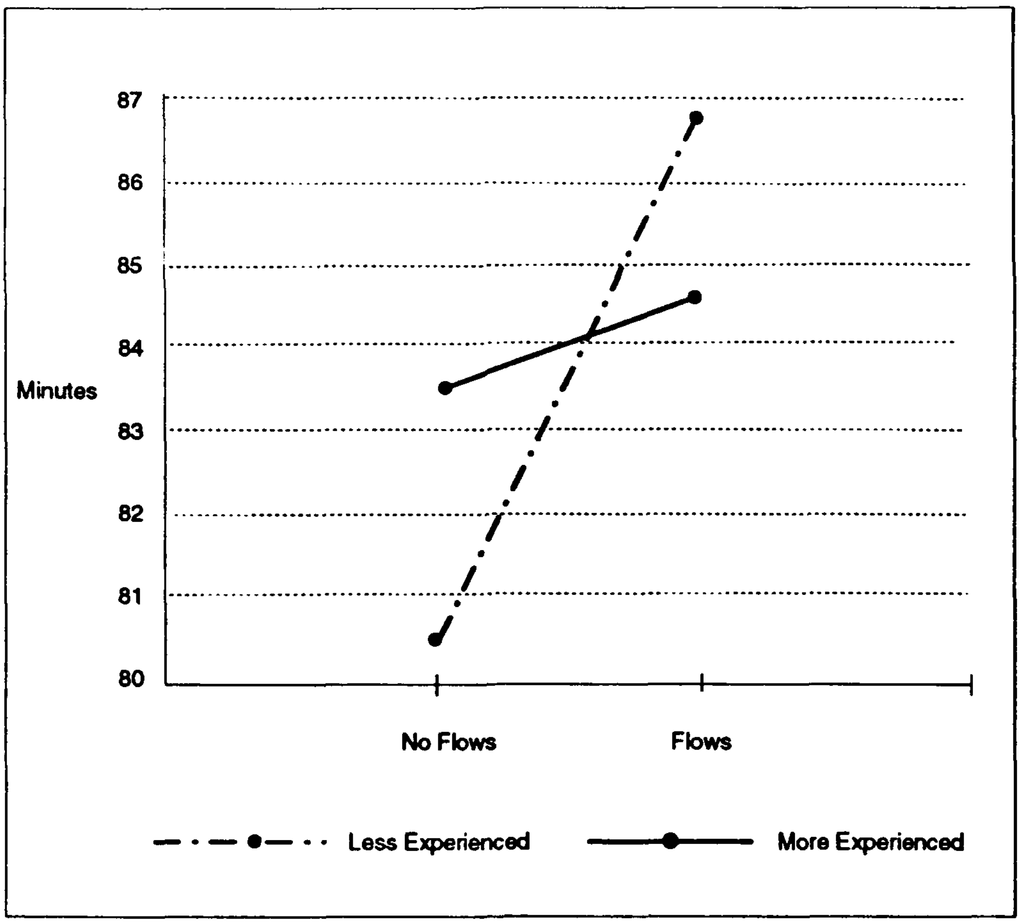

Figure 21. Effect of Experience and Availability of High Abstraction Displays on Maintenance Performance (Experiment 4)

Time to Solution. Experience level was the only independent variable that had a significant effect on time to solution. The high experience subjects solved the problems 38% faster than the low experience subjects (4.8 versus 6.6 minutes, $F(1,12) = 11.82$, $p < .01$).

Diagnostic Errors. The number of diagnostic errors showed no significant variance with any of the three main effects. However, there was a significant three-way interaction among the effects (see Figure 22, $F(1,12) = 4.94$, $p < .05$). The low experience group that did not receive the "how to use" training performed much worse on the trials with high abstraction displays available (incorrect diagnosis rate of 31%) than without the high abstraction displays (incorrect diagnosis rate of 7%). The experienced group that did not receive the "how to use" training improved on trials with high abstraction displays available (incorrect diagnosis rate of 17% without and 6% with high abstraction displays). There was less variation in the other two groups (high and low experience subjects that received the "how to use" training), with all conditions ranging from 13% to 20%.

Display Usage. For this analysis, "use" and "usage" refer to the percentage of time that a particular type of display was used in a trial. In the differences reported here, a 10% difference is approximately 30 seconds.

The within-subjects variable (availability of high abstraction displays) had a significant, but predictable effect on the usage of the abstraction levels. The use of medium abstraction displays was significantly lower on the trials with high abstraction displays available (78% versus 93%, $F(1,12) = 19.32$, $p < .005$). The use of low abstraction displays reflected a similar tendency, although the difference was not as large (5% versus 8%) nor was it statistically significant.

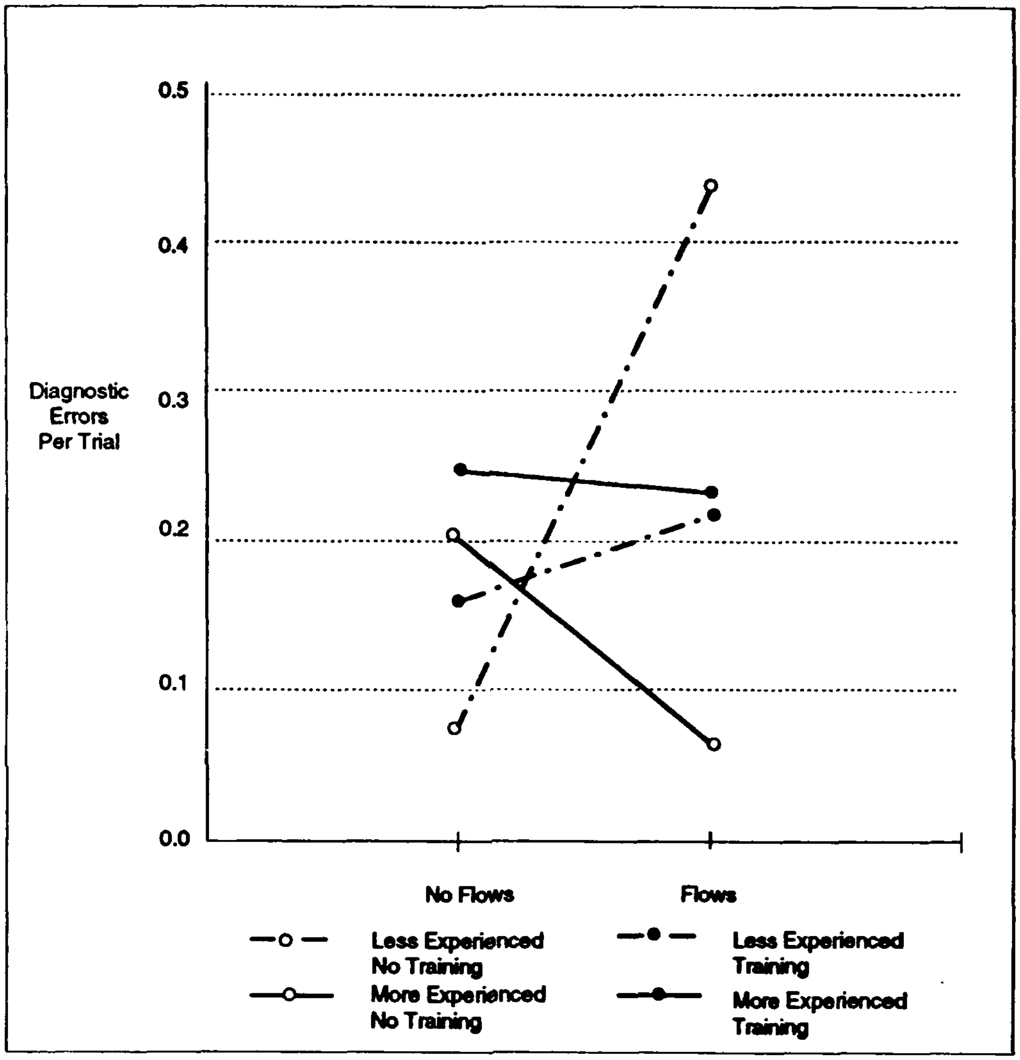

Figure 22. Effect of Experience, "How to Use" Training, and Availability of High Abstraction Displays on Number of Diagnostic Errors Per Trial

Experience level and type of training showed no effect on the usage of the high abstraction displays. However, the group that received the "how to use" training used the medium abstraction displays less (80% versus 91%, $F(1,12) = 8.64$, $p < .05$) and the low abstraction displays more (9% versus 3%, $F(1,12) = 20.39$, $p < .005$) than did the other group. Similarly, the low experience group used

the medium abstraction displays less (81% versus 90%, $F(1,12) = 7.24$, $p < .05$) and the low abstraction displays more (9% versus 4%, $F(1,12) = 17.09$, $p < .005$) than did the high experience group.

In terms of the number of transitions among displays, the low experience group made 72% more transitions among displays than did the high experience group (mean = 6.2 versus 3.6, $F(1,12) = 14.51$, $p < .01$). The group that received the "how to use" training made 42% more transitions than did the other group (mean = 5.7 versus 4, $F(1,12) = 13.42$, $p < .01$).

None of the three variables had a significant effect on the usage of aggregation levels.

Display Errors. None of the three variables had a significant effect on the number of display access errors or on the time spent viewing inappropriate displays. Less than 5% of the display accesses were to inappropriate displays, and only 3.6% of the time (or about 15 seconds per trial) was spent viewing inappropriate displays.

F. Discussion - Experiment Four

Of the independent variables, only experience level and training type exhibited main effects on the dependent measures. The measures affected were time to solution, number of display transitions, and usage of the medium and low abstraction displays.

Time to solution was affected by experience level; the more experienced subjects solved the problems faster than the less experienced subjects. While this result is intuitively pleasing, it is not too surprising. It is not possible to say how this performance compares with either group using traditional paper-based displays. Neither the availability of the high abstraction displays nor the "how to use" training significantly affected the time that it took to solve the problems.

As a main effect, differences in experience level were associated with differences in the number of transitions among displays and the percentage of time spent in the medium and low abstraction levels. In comparison with the low experience group, the high experience group had fewer transitions among displays, more use of the medium abstraction displays, and less use of the low abstraction displays. These same results were seen in the comparison of the two training groups. The group that did not receive the "how to use" training also had fewer transitions among displays, more use of the medium abstraction displays, and less use of the low abstraction displays. These two groups seemed to solve the problems with only a few display selections and mostly using medium abstraction displays.

The most probable explanation for this difference between the experience levels is that the high experience subjects did not need the low abstraction displays (location diagrams) to locate the test and replacement points. Therefore, they did not need to transition to the location diagrams for each test or repair; whereas the less experienced subjects, who may be less familiar with the system, needed to view the location diagrams in order to identify the test and replacement points. This would account for both the increased number of transitions as well as the increased use of the low abstraction displays.

The difference between the training groups is more difficult to explain. The behavior of the group that received the "how to use" training was similar to that of the low experience group (i.e., more display transitions and a broader mix of abstraction levels). A plausible explanation is that the inexperienced subjects and the subjects that received the "how to use" training were more confused or uncertain about what displays were needed. However, the lack of significant differences in the analysis of the use of inappropriate displays argues against this. It is more plausible that the portion of the training that compared the usefulness of the different

abstraction levels in different situations made the subjects more comfortable moving among displays and using a wider variety of displays.

Neither of these between-subjects variables (experience or training type) showed a significant effect on the primary dependent measure (simulated maintenance time); so we cannot claim that training that promotes using a wider variety of displays necessarily improves overall performance. And since the lower time to solution was associated with only experience level and not training type, we cannot infer that this type of training helps solve problems more quickly. It does appear that this type of training can foster broader, more flexible use of a display system.

There were only two significant effects on simulated maintenance performance, and they resulted from interactions among the independent variables. First, the "how to use" training did not improve the performance of the experienced subjects, but it did improve the less experienced subjects. The less experienced group as a whole exhibited broader use of the display system. This suggests that simply using novel graphics is inadequate for improving the performance of less experienced maintainers; it is necessary to train them on the strategies that are useful with the displays. More experienced subjects may already use these strategies or have others that are equally useful, since the "how to use" training did not affect their performance.

Second, these results hinted that providing the high abstraction displays do not improve maintenance performance in troubleshooting, and that the low experienced subjects perform more poorly with the high abstraction displays available. An analysis of the components of the maintenance performance measure (time to solution and time to perform tests and repairs) revealed no single component that explained this two-way interaction (experience by flow availability). The three-way interaction on the number of diagnostic errors (incorrect repairs) is

the only significant parallel. In this result, the low experience group that did not receive the "how to use" training had an incorrect diagnosis rate of 31% on trials with the high abstraction displays available and only 7% without them available. There was no significant difference in this group's usage of the high abstraction displays, when compared with the other three groups. This could imply that the high abstraction displays did not enable these subjects to reach the correct diagnoses. Further investigation of this was planned for the final experiment.

G. Experiment Five

The previous experiment indicated that maintainers' performance is influenced by the availability of high abstraction displays, and that the influence differs depending on the experience level of the maintainer. The next step in our research was to investigate how performance is affected by changing the display mix available to the maintainer.

Of greatest interest is the decrease in performance resulting from providing the high abstraction displays to the low experienced subjects in Experiment Four. However, we were unable to obtain the right mix of subjects to use experience level as a between-subjects variable since we had nearly exhausted the available population of maintainers through prior experiments. Consequently, we pooled the available maintainers and proceeded with a within-subjects design.

1. Experimental Design

As in experiment 4, each trial was a troubleshooting task. In each of six trials, the subject was given a failure symptom resulting from a single failure, and the subject used the maintenance information system to identify the failure. The trials were conducted in the same manner as in experiment 4.

Independent Variable. The experiment used a one factor design with one within-subjects variable, *display condition.* This within-subjects variable was intended to study the effect of different combinations of abstraction and aggregation levels on maintenance performance. Each subject was given two trials under three different display conditions (a total of six trials) providing different combinations of availability of the medium abstraction/medium aggregation displays (function-oriented schematics) and the high abstraction/medium aggregation displays (function-oriented flow diagrams). Figure 23 illustrates these three conditions. The availability was balanced across trials to reduce bias resulting from differences in difficulty among the trials. The subjects were informed about the availability of the different types of displays at the beginning of each trial.

Based on the observation of subjects in prior experiments, it was expected that performance would be best on display condition 1 (with the function-oriented schematics) and worst on the baseline condition (with only the device-oriented schematics). It was less clear how display condition 2 (with the function-oriented flow diagrams) might affect performance. An important issue in this experiment is whether the high abstraction/medium aggregation displays (functional flow diagrams) could improve maintenance performance as much as the medium abstraction/medium aggregation displays (functional schematics). The benefits of this comparison will be discussed later.

Dependent Measures. The measures in this experiment were the same as in experiment 4 (simulated maintenance performance, time to solution, diagnostic errors, and display usage), with the exception that display errors were not analyzed, since the measure produced no practically significant results in experiments 3 and 4.

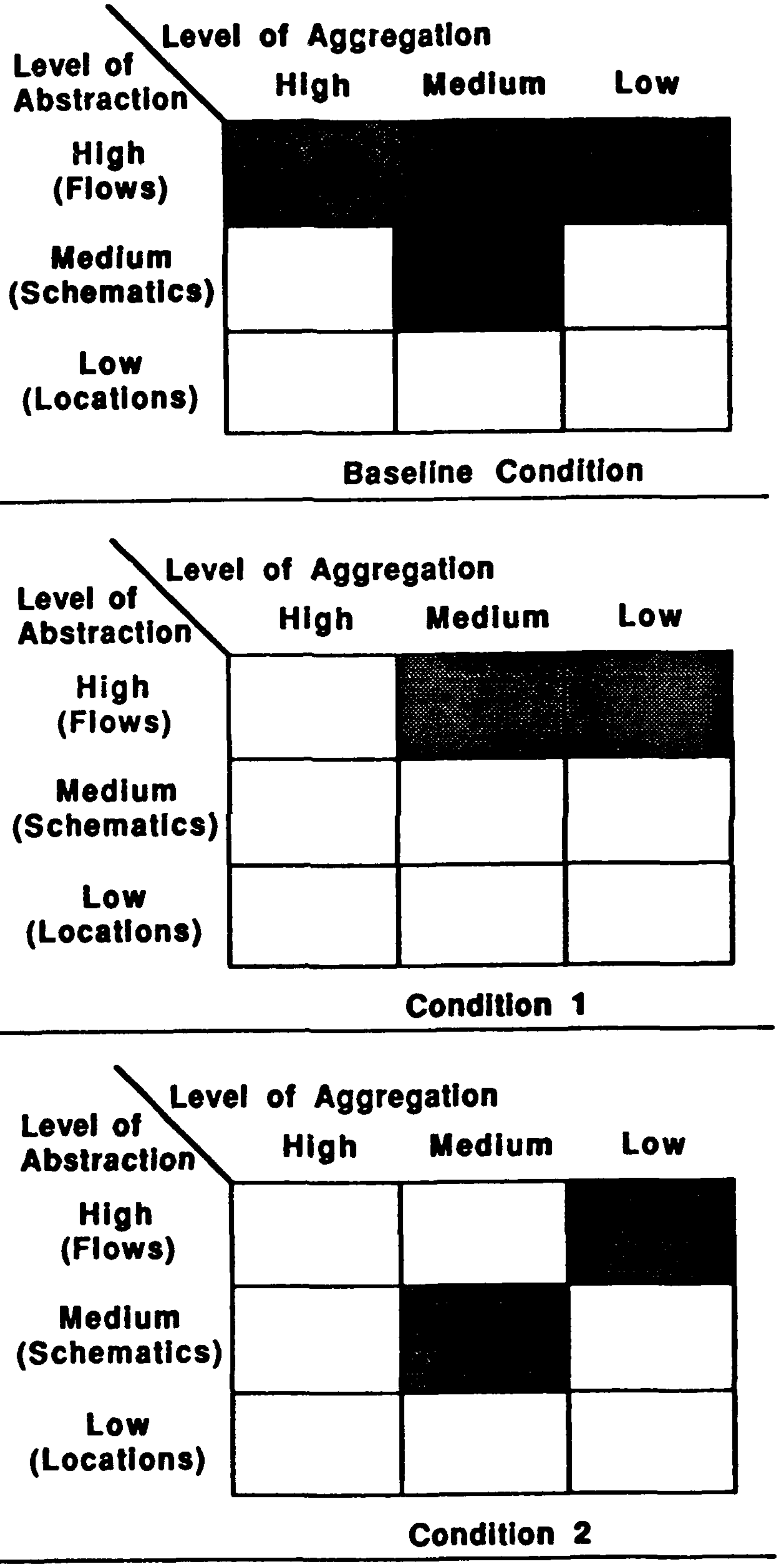

Figure 23. Experimental Conditions for Experiment 5

2. Subjects

As in the earlier experiments, the subjects were actual maintenance personnel, trained and experienced with the SH-3 blade fold system. The ten subjects were E3's to E8's with 5 to 14 years of maintenance experience (median = 8.13, mean = 8.84, standard deviation = 3.5 years).

Before the trials, each subject received the same training and practice, which consisted of approximately 45 minutes of individual use of the display system. The training included the "how to use" training from Experiment Four.

3. Results

The data from each trial were standardized as z-scores, and an ANOVA was performed to determine the effects of display condition on the dependent measures. Nine trials (of a total of 60) that were not completed in the allotted time were excluded from the analysis. Of these nine trials, seven were under the baseline display condition and two under display condition 2. Also excluded were five trials in which the main performance measure (simulated maintenance time) was outside two standard deviations of the mean for the trial. Of these five trials, two were under the baseline display condition and three under display condition 1.

Maintenance Performance. Maintenance performance showed a significant difference based on display condition ($F(2,18) = 3.85$, $p < .05$) -- see Figure 24. The subjects' performance under display condition 2 was significantly better than under the baseline condition ($F(1,9) = 5.54$, $p < .05$). While the average performance under display condition 1 was better than the baseline condition, the difference was not significant ($F(1,9) = 4.40$, $p = .065$). The difference between conditions 1 and 2 was not significant.

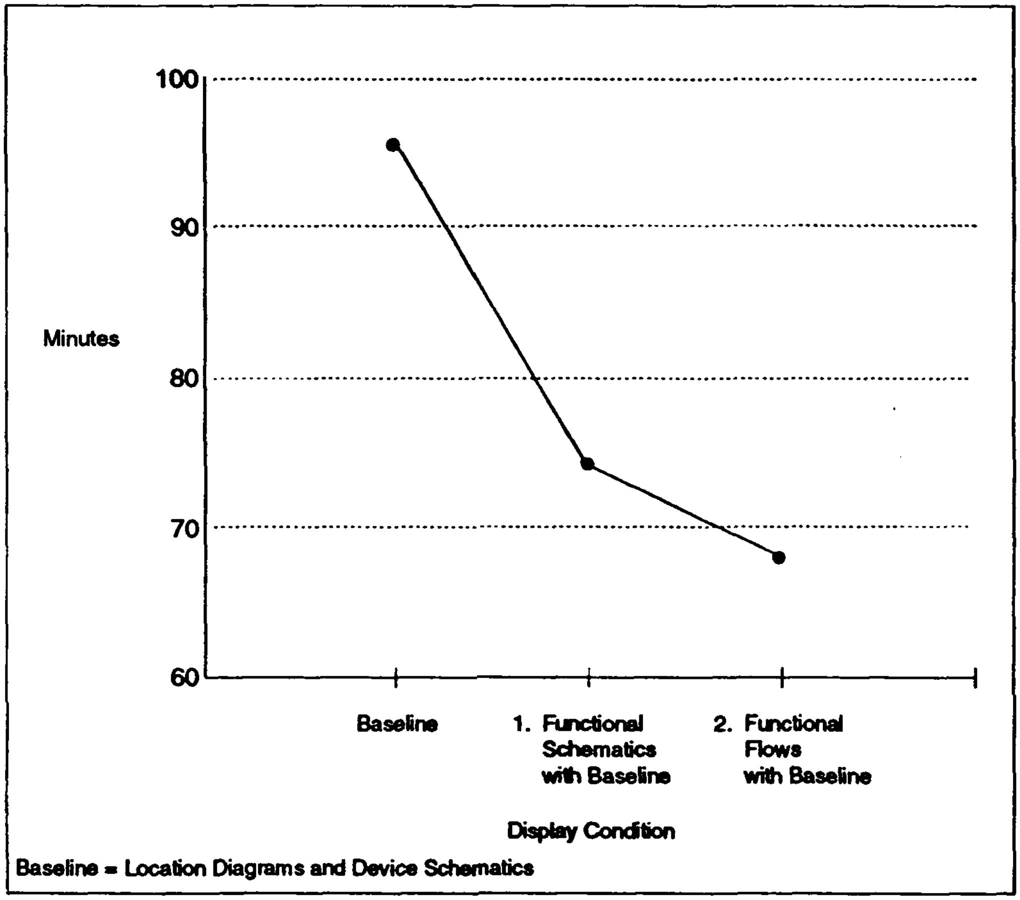

Figure 24. Effect of Display Availability on Maintenance Performance (Experiment 5)

Time to Solution. Display condition also had a significant effect on the total time to solution ($F(2,18) = 7.46$, $p < .01$). The baseline display condition resulted in significantly longer times to solution (mean = 9.5 minutes) than did display condition 1 (5.4 minutes, $F(1,9) = 17{,}56$, $p < .01$) or display condition 2 (6.4 minutes, $F(1,9) = 7.98$, $p < .05$). The difference between display condition 1 and 2 was not significant.

Diagnostic Errors. Display condition had no significant effect on diagnostic errors. The mean error rate was less than 10%.

Display Usage. The usage of the abstraction levels followed predictable patterns because of the display conditions selected. For high abstraction displays, display condition 2 had the highest use and condition 1 had the lowest use (28% versus 4%, $F(1,9) = 19.4$, $p < .005$). High abstraction displays were not available in the baseline display condition. The medium abstraction displays were used less in condition 2 (69%) than in condition 1 (91%, $F(1,9) = 12.34$, $p < .01$) or the baseline condition (96%, $F(1,9) = 22.87$, $p < .01$). Low abstraction displays were used less than 5% of the time for all display conditions, and there were no significant differences.

The usage of the high aggregation displays was between 8% and 11% for all three conditions, and the differences were not significant. The medium aggregation displays were used 74% in display condition 1, 23% of the time in display condition 2, and 1% in the baseline display condition. These differences were all significant at the $p < .005$ level. The low aggregation displays were used 17% in display condition 1, 70% of the time in display condition 2, and 88% in the baseline display condition. These differences were all significant at the $p < .05$ level.

There was also a significant difference among the display conditions in terms of the number of display transitions within a trial ($F(2,18) = 12.4$, $p < .001$). Display condition 1 had the fewest transitions per trial (mean = 3.7). Display condition 2 and the baseline display condition had significantly more (11.8 for the baseline, $F(1,9) = 28.3$, $p < .001$; and 8.6 for display condition 2, $F(1,9) = 6.9$, $p < .05$). The difference in the number of transitions between display condition 2 and the baseline was not statistically significant.

H. Discussion - Experiment Five

The three display conditions in this experiment (baseline, condition 1, and condition 2) represent different ways in which particular abstractions and aggregations can be combined in a display system.

The baseline condition, the minimum display set, was used to provide a baseline of comparison for the other two display conditions. This baseline display set, which was augmented with additional displays in conditions 1 and 2, was sufficient to solve all of the problems presented in the trials. The baseline condition consisted of a full set of low abstraction displays (location diagrams) and a set of medium abstraction/low aggregation displays (device-oriented schematic diagrams). These schematics portray the same level of detail as the paper-based displays currently used by these maintainers. However, they have been partitioned and redrawn to fit on the computer screen. In this partitioning, components that make up a subassembly were placed on the same display; if a subassembly contained too many components to fit on one display, then more displays were designed, which taken together encompass the entire subassembly. This is the same method used in the paper-based documents to divide the overall system schematic into several pages of schematics.

Condition 1 added displays consisting of a different aggregation method in the medium abstraction level. These medium abstraction/medium aggregation displays (function-oriented schematics) were used the most in all of the previous experiments. These displays are especially useful for troubleshooting, since they are aggregated by function and the expression of troubleshooting problems is usually as the failure of some function. Furthermore, for most functions, a single display contains all of the necessary information to understand how a circuit operates <u>and</u> to identify test points during troubleshooting. In addition, condition 1

had one high abstraction/high aggregation display (a flow diagram), which served as an index into the medium abstraction/medium aggregation displays.

Instead of function-oriented schematics, condition 2 added displays that were different in both abstraction and aggregation when compared to the baseline. These high abstraction/medium aggregation displays (function-oriented flow diagrams) were also aggregated by the function of the circuit, and they, too, contained all of the necessary information to understand how a circuit operated. These diagrams could also be used to identify test points in a very general sense (e.g., between components A and B), but they did not contain sufficient information to identify specific test points (e.g., pin 1 on component A). The schematic diagrams had to be used for that information. These flow diagrams employed simpler graphics (boxes and lines) than the schematics, enabling slightly higher aggregation levels and additional coding that was not always possible with the function-oriented schematics (e.g., left-to-right and top-to-bottom flow of signals). Condition 1 also had one high abstraction/high aggregation display, which served as an index into the high abstraction/medium aggregation displays.

The significant improvement in maintenance performance and time to solution under conditions 1 and 2 (in comparison to the baseline) recalls a similar result in Experiment One using paper-based displays. Providing displays with a variety of levels of aggregation and abstraction improved performance. In particular, performance was improved by 1) adding a level of aggregation within an existing level of abstraction in condition 1, or 2) by adding an additional level of abstraction in condition 2.

Due to the limited size of the available subject pool, we were unable to further investigate whether providing high abstraction displays to inexperienced maintainers increases errors in troubleshooting. The overall rate of incorrect diagnosis was lower in this experiment (10% versus 13-20% for subjects with "how to use" training).

For the relatively experienced subjects in this experiment, there was no difference in the diagnostic errors based on display condition.

V. RESULTS FROM THE OPINIONNAIRES

At the conclusion of each trial in each of the five experiments, the subjects completed an opinionnaire. The results generally supported the usefulness of this approach to designing computer-based graphics for maintenance (see Figure 25). Two responses are particularly worth noting. The subjects in all experiments indicated that they experience only moderate confusion or frustration with the standard maintenance documentation. However, 89% of the subjects indicated that this new approach would provide improvement in supporting maintenance activities. Sixty-four percent indicated it would provide *much* improvement.

At least two separate issues may contribute to the generally favorable response to this maintenance information system. One, which is the focus of this research, is that the novel display formats are more appropriate for maintenance activities than the traditional display formats. However, another possible reason is that the flexible display access provided by a computer-based maintenance information system is much more convenient to use than traditional paper-based manuals. It is also likely that both the novel displays and the easy access are necessary for an effective system. This research is addressing the first issue. The results from the current industry-wide emphasis on hypertext applications will certainly address the second issue. Given the inevitable movement toward electronic maintenance documentation, these opinions from professional maintainers reinforce the validity of this approach to providing computer-based graphics for maintenance activities.

	Good	OK	Bad
1. Adequacy of display sizes for displaying information	83.3%	12.5%	4.2%
2. Adequacy of spacing among elements in the displayed information (i.e., lack of clutter/crowding)	93.9%	6.1%	0.0%
3. Adequacy of arrangement of displayed information elements (e.g., schematics, block diagrams)	87.8%	12.2%	0.0%
4. Adequacy of contrast between information displayed and background	87.8%	10.2%	2.0%
5. Adequacy of resolution and clarity of information elements displayed	81.6%	18.4%	0.0%
6. Adequacy of detail of information elements displayed	81.6%	18.4%	0.0%
7. Adequacy of legibility of displayed letters and words	83.7%	16.3%	0.0%
8. Adequacy of the organization and arrangement of maintenance information	89.8%	10.2%	0.0%
9. Adequacy of maintenance diagrams for use in troubleshooting	87.8%	10.2%	2.0%
10. Ease of finding different general types of maintenance information	81.6%	18.4%	0.0%
11. Ease of finding specific information within a particular type of maintenance information	81.6%	16.4%	2.0%
12. Adequacy of maintenance information supporting troubleshooting	73.5%	22.4%	4.1%

Figure 25. Opinionnaire Responses (Based on Smillie et al., 1988)

	Much	Some	Little	None
13. Amount of confusion or frustration you currently experience in obtaining needed maintenance information from standard maintenance materials	2.1%	68.7%	16.7%	12.5%
14. Amount of improvement in the overall organization and arrangement of maintenance information that this new material provides	55.1%	32.6%	8.2%	4.1%
15. Amount of improvement in the presentation of maintenance information that this new material provides	42.9%	40.8%	10.2%	6.1%
16. Amount of improvement in the overall completeness, accuracy, and applicability of maintenance information that this new material provides	45.0%	36.7%	16.3%	2.0%
17. Amount of improvement in supporting maintenance on the SH3 bladefold that this new material provides	63.8%	25.5%	8.5%	2.2%
18. Amount of improvement in supporting maintenance on the SH3 bladefold that the flow diagrams provide	64%	20%	8.0%	8.0%
19. Amount of improvement in supporting maintenance on the SH3 bladefold that the function schematic diagrams provide	57.8%	34.6%	3.8%	3.8%
20. Amount of improvement in supporting maintenance on the SH3 bladefold that the device schematic diagrams provide	57.7%	30.8%	3.8%	7.7%
21. Amount of improvement in supporting maintenance on the SH3 bladefold that location diagrams provide	65.4%	19.2%	7.7%	7.7%

Figure 25. Opinionnaire Responses (cont.)

VI. GENERAL DISCUSSION AND GUIDELINES

Designing displays for computer-based access to large graphical databases requires consideration of the nature of the system and the user's tasks as well as the user's model of the system and tasks. These experiments have investigated some of these considerations, and this section summarizes some of the lessons learned in developing the maintenance information system and observing over fifty Navy maintenance personnel using the system. Figure 26 presents a summary of the five experiments, and Figure 27 lists twelve guidelines developed as a result of this work. These are general guidelines, and they constitute more of a checklist rather than a prescriptive process.

1. Identify potential roles for abstractions and aggregations. One reason for providing displays with different abstractions and aggregations can be to <u>aid</u> or <u>train</u> users on the nature of the system or their tasks within the system. Another possible use of abstractions and aggregations is to provide mechanisms for <u>navigation</u> and <u>retrieval</u> within a graphical database.

2. If possible, investigate existing documentation for useful abstractions and aggregations. For existing systems, the documentation may provide valuable information on abstractions and aggregations with which the user is already acquainted. In the SH-3 example application discussed earlier, the location diagrams and device-oriented schematic diagrams were derived from existing documentation. The notion for the function-oriented schematics came from sketches used in maintenance training classes.

xExperiment	Variables	Measures	Subjects	Results
1	• Experience • Displays	Number of displays	6	• Fewer displays used with experimental material.
2 3	• Experience • Task type	Time on abstraction & aggregation levels	10 13	• "Thinking" tasks require higher abstraction. • "Doing" tasks require lower abstraction. • On "thinking" tasks, experienced S use high abstraction displays more than inexperienced S. • Aggregation use shifts to lower aggregation as task progresses.
4	• Experience • Training content • Displays	• Time to solution • Simulated maintenance performance • Time on abstraction & aggregation	16	• "How to use" training improved performance of inexperienced S. • Inexperienced S performed more poorly with high abstraction displays.
5	• Displays	• Time to solution • Simulated maintenance performance	10	• Performance improved by increasing either abstraction or aggregation over baseline displays.

Figure 26. Summary of Experiments

1. Identify potential roles for abstractions and aggregations.
2. If possible, investigate existing documentation for useful abstractions and aggregations.
3. Provide a variety of abstraction levels appropriate for the user's tasks.
4. Review the phenomena underlying the function of the system to identify potential abstractions.
5. Review the user's tasks to identify potential abstractions.
6. Provide a range of aggregation levels appropriate for the user's tasks.
7. Review the user's tasks to identify potential aggregations.
8. Review the user's experience and background to ensure that the abstractions will be meaningful.
9. Train the users, especially those with limited experience, on how and when to use the various abstractions and aggregations.
10. Review the target display technology to ensure that the aggregations are achievable.
11. Use abstractions and aggregations in creating documentation.
12. Evaluate logistical issues with maintaining the display system itself.

Figure 27. Abstraction/Aggregation Guidelines

3. Provide a variety of abstraction levels appropriate for the user's tasks. Experiments Two and Three supported our hypothesis on the relationship between task type and the appropriate abstraction levels for displays. Maintainers tended to use displays with higher abstraction levels more when performing tasks that have a larger "thinking" component than when performing tasks with a larger "doing" component. Similarly, maintainers tended to use displays with lower abstraction levels more when

performing tasks that have a larger "doing" component than when performing tasks with a larger "thinking" component.

4. Review the phenomena underlying the function of the system to identify potential abstractions. One way to identify potential abstractions is to look below the physical activities of a system to the functions that the activities accomplish. For example, heat transfer and feedback control are conceptually more abstract than the mechanisms that actually implement the activity of heating, cooling, or moderating. In the SH-3 blade fold system, the electrical relay interlocks and hydraulic sequencing valves are two examples of this. In this system, the electrical subsystem consists of interlocking relay logic that ensures the safety of the equipment during all phases of operation. The electrical interlocks were the basis for the electrical flow diagrams that were designed to show how the permissives operated. The hydraulic sequencing valves inspired a variation on the flow diagrams that illustrated how hydraulic pressure was sequenced through the system during operation.

5. Review the user's tasks to identify potential abstractions. Tasks also occur at different levels of abstraction, and useful abstractions may be identified by looking below the physical activities of the human. For example, different abstractions may be useful for troubleshooting based on topology of a circuit and troubleshooting based on symptom-failure mapping. Procedural and non-procedural tasks are also likely to benefit from different abstractions.

As discussed earlier, the user's model of the system and tasks can be influenced by the selection of abstractions to represent the system and

tasks. Conversely, the user's models can be used (if they are "correct") to identify useful abstractions for presenting information that supports a task or explains the system.

In the second and third experiments, the "following procedures" tasks used fully proceduralized job performance aids, which contained task models for troubleshooting certain failures. These task models are based on binary decision trees, which are most useful for novice maintainers. However, locating the test points for these binary decisions requires a rather extensive knowledge of the layout of the system within the helicopter, knowledge which novice maintainers are unlikely to command. This binary task model (itself a task abstraction) can be linked with location diagrams (low abstraction displays of the system) in a maintenance information system to provide effective support for novice maintainers.

6. Provide a range of aggregation levels appropriate for the user's tasks. The results of Experiments Two and Three illustrate the need for a range of aggregation levels. In all task types, the maintainers began with the high aggregation displays and finished with the low aggregation displays. This is also supported by the differences in maintenance performance in Experiment Five. In the baseline display condition of Experiment Five, the schematics were provided only in the low aggregation level (device-oriented schematics). In display condition 1, the medium aggregation schematics (function-oriented schematics) were made available, and the maintainers' performance improved significantly.

7. Review the user's tasks to identify potential aggregations. This basic tenet of display design bears mentioning here. The notion of aggregating information requirements to support a task can be extended to the display of static graphical information. When a display must be divided to fit within the display space, this guidance can be used to decide how to divide it.

   In the example system, we used this guidance to decide how to partition the device-oriented schematics for the relay panel. Since our user's tasks included troubleshooting and circuit tracing, we sought to minimize the number of links among the schematics of the relay panel. In doing this, we minimized the number of display changes necessary to trace through a given path in the relay logic. For a different system, aggregating by physical proximity (for electromagnetic interference problems) or components using common resources (e.g., power or space for resource allocation problems) may be more important.

8. Review the user's experience and background to ensure that the abstractions will be meaningful. One result from Experiment Four indicated that the low experience group of maintainers performed more poorly when the high abstraction displays were available. While the number of subjects in the experiment is too low to say with certainty, it is reasonable to suggest that that group of subjects was unable to effectively use the high abstraction displays or that the displays did not enable them to reach valid conclusions. Also, the low experience subjects tended to favor the medium abstraction displays over the high abstraction displays on the problem solving tasks in Experiment Three. Once again, it is reasonable to suggest that they found the flow diagrams

(high abstraction displays) more difficult to use and preferred the more familiar schematic diagrams (medium abstraction displays).

The abstractions chosen should be evaluated for the full range of anticipated users; and if some abstractions are found to be inappropriate for some users, perhaps another abstraction could be used. Otherwise, the users could be trained or aided to improve their performance with the displays.

9. Train the users, especially those with limited experience, on how and when to use the various abstractions and aggregations. The results from Experiment Four suggested the importance of training inexperienced maintainers on how and when to use the various aggregations and abstractions. Merely defining the different types of displays and explaining the symbology of the displays and the relationships among displays may be inadequate. It is important to define for the maintainer the types of situations that the different kinds of displays were designed to support and how to use the displays in those situations. Comparisons of trying to use different kinds of displays in a given situation seemed to be a particularly effective communication tool in Experiment Four.

10. Review the target display technology to ensure that the aggregations are achievable. This research is motivated by the inability of current and anticipated display technology to contain the full range of desired aggregation levels for displays. While this approach seeks to ease the problem through the use of different aggregations and abstractions, the constraint remains. There is likely to be a tradeoff among size, weight, and cost that will affect display design.

11. Use abstractions and aggregations in creating documentation. The abstractions and aggregations should be used in creating maintenance and training documentation for a system. This applies to documentation that complements the computer-based information system as well as stand-alone documents.

12. Evaluate logistical issues with maintaining the display system itself. Some graphical displays are easier to create and change than others. For example, in the maintenance information system developed for this research, the flow diagrams were much easier to create and update than the schematic diagrams. There is much more flexibility in the layout of block diagrams than in the layout of schematic diagrams, but this flexibility comes with decreased levels of detail.

    Experiment Five suggested that overall maintenance performance when using function-oriented flow diagrams is not different than the performance when using function-oriented schematic diagrams. This indicates that for these tasks and types of subjects these displays are redundant. In this situation, for logistical support of the display system, the flow diagrams should be chosen over the functional schematics, because of the ease of display changes and updates to the flow diagrams. However, even in this limited case, this conclusion should be tempered by the fact that Experiment Five was not able to evaluate the effects of experience level, and Experiment Four indicated that low experience maintenance personnel may not perform well using high abstraction displays.

Nonetheless, if other issues are equal, the maintainability of the display system itself is an issue to consider in selecting abstraction and aggregation levels.

VI. CONCLUSIONS

The focus of this research was to develop principles that can be used in designing computer-based displays of graphical maintenance information, which is traditionally contained in large, high resolution, paper-based drawings. Related research on many different approaches to this problem were discussed earlier. This research is unique in its explicit use of the display aggregation/abstraction space described by Rasmussen (1986), as illustrated earlier in Figure 9.

In the experiments performed under this research program, experienced maintainers used a computer-based display system to gather information that they would use in performing actual maintenance tasks. Furthermore, the maintainers were trained and experienced in the domain (i.e., the SH-3 blade fold system). Because of this, these results are particularly encouraging for the application of these concepts to the development of computer-based maintenance information systems.

These experiments have shown that:

1. Displays designed using principles of aggregation and abstraction can improve maintenance performance over that obtained using displays designed using the principles in current paper-based documents.

2. The maintainers' choice of display abstraction level is influenced by both their experience level and the maintenance task at hand.

3. The maintainers' choice of display aggregation level depends on whether they are in the early or late stages of a maintenance task.

4. Training on "how to use" the abstractions and aggregations during troubleshooting can help improve the performance of inexperienced maintainers.

5. High abstraction displays may not improve maintenance performance, especially for inexperienced maintainers without "how to use" training.

6. The maintainers demonstrated an apparent preference for the novel display formats over the traditional formats, probably due to the fact that the novel displays were more suitable for their tasks.

This report has discussed a variety of display design issues that are particularly problematic in complex systems. A model-based framework for pursuing these issues was described. This framework is, admittedly, quite ambitious as well as very preliminary. Nevertheless, it can provide important direction for research as well as design practices.

Several hypotheses emerging from this framework were tested in the five experiments whose results were discussed here. In general, these hypotheses were supported, although not completely. Of course, this does not, by any means validate the overall framework as "correct." It does, however, show that this approach is interesting and useful.

ACKNOWLEDGMENT

This report is based on research supported by the Office of Naval Research and the Naval Training Systems Center under Contract No. N00014-89-C-0047. The authors gratefully acknowledge the participation and contributions to this effort by Dr. William J. Cody and Ms. Alana M. Anoskey of Search Technology. The authors would also like to thank AEC Charles Gaylord and AE1 Ron Barry of HS WING ONE, Jacksonville Naval Air Station for providing domain expertise and for their efforts in arranging for the subjects in these experiments.

REFERENCES

Aerospace Industries Association Pubs-120. 1989. *Guide to Simplified Graphics.*

Bury, K.F., Boyle, J.M., Evey, J.M., and Neal, A.S. 1982. Windowing versus scrolling on a visual display terminal. *Human Factors*, Vol. 24, pp. 385-394.

Duchnicky, R.L. and Kolers, P.A. 1983. Readability of text scrolled on visual display terminals as a function of window size. *Human Factors*, Vol. 25, pp. 683-692.

Duncan, K.D. 1982. Display ergonomics and fault location. *Proceedings of the Second European Annual Conference on Human Decision Making and Manual Control*, pp. 319-330.

Flanagan, D.P., Blue, D.V., Giacaglia, R.A., Lenorovitz, D.R., and Stanke, E.C. 1987. *Rapid Intelligent Prototyping Laboratory (RIPL) Architecture and Use* (Tech. Rept. TR-87-004). Wright-Patterson Air Force Base, OH: AAMRL.

Frey, P.R. 1989. *Knowledge Aided Display Design (KADD): A Decision Support System for Display Design.* Hampton, VA: National Aeronautics and Space Administration.

Frey, P.R. and Wiederholt, B.J. 1986. KADD - An environment for interactive knowledge-aided display design. *Proceedings of the 1986 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1472-1477.

Frey, P.R., Sides, W.H., Hunt, R.M., and Rouse, W.B. 1984. *Computer-generated Display System Guidelines, Volume 1: Display Design* (Tech. Rept. NP-3701, Vol. 1). Palo Alto, CA: Electric Power Research Institute.

Furnas, G.W. 1986. Generalized fisheye views. *CHI '86 Proceedings*, pp. 16-23.

Garris, R.D., Mulligan, C.P., Ricci, K.E., Dwyer, D.J., McCallum, G.A., and Moskal, P. 1990a. *Effects of Graphics Detail Parameters on Locator Task*

Performance for Computer-based Training (NAVTRASYSCEN TR-88-020). Orlando, FL: Naval Training Systems Center.

Garris, R.D., Mulligan, C.P., Ricci, K.E., and McCallum, G.A. 1990b. *Effects of Graphics Detail Parameters on Disassembly/assembly Task Performance for Computer-based Training* (NAVTRASYSCEN TR-90-020). Orlando, FL: Naval Training Systems Center.

Geiser, G. and Schumacher, W. 1976. Parallel vs. serial instrumentation for multi-variable manual control in control rooms. In T.B. Sheridan and G. Johannsen (Eds.), *Monitoring Behavior and Supervisory Control*, New York: Plenum Press.

Glushko, R.J. 1989. Hypermedia: Visions of grandeur? *UNIX Review*, Vol. 8, pp. 70-80.

Henneman, R.L. and Rouse, W.B. 1984a. Human performance in monitoring and controlling hierarchical large-scale systems. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-14, pp. 184-191.

Henneman, R.L. and Rouse, W.B. 1984b. Assessing the complexity of a large scale system: Measures of system structure and human strategy. *Proceedings of the 1984 IEEE Systems, Man, and Cybernetics Conference*, pp. 68-72.

Hunt, R.M. and Frey, P.R. 1987. Knowledge-aided display design (KADD) system: An evaluation. *Proceedings of the 31st Annual Meeting of the Human Factors Society*, pp. 536-540.

Mitchell, C.M. and Miller, R.A. 1983. Design strategies for computer-based information display in real-time control systems. *Human Factors*, Vol. 25, pp. 353-369..

Morehead, D.R. and Rouse, W.B. 1983. Human-computer interaction in information seeking tasks. *Information Processing and Management*, 19, pp. 243-253.

Morris, N.M., and Rouse, W.B. 1985a. Review and evaluation of empirical research in troubleshooting. *Human Factors*, 27, pp. 503-530.

Morris, N.M. and Rouse, W.B. 1985b. The effects of type of knowledge upon human problem solving in a process control task. *IEEE Transactions on Systems, Man, and Cybemetics,* 15, pp. 698-707.

Perlman, G. and Moorhead, T. 1988. *Navitext SAM: Software user manual.* Brookline, MA: Northern Lights Software Corporation.

Rasmussen, J. 1986. *Information Processing and Human-machine Interaction: An Approach to Cognitive Engineering.* New York: North-Holland.

Ricci, K.E., Garris, R.D., and McCallum, G.A. 1990. *Effects of Graphic Detail Parameters on Procedural Task Performance for Computer-based Training* (NAVTRASYSCEN TR89-023). Orlando, FL: Naval Training Systems Center.

Rouse, S.H. and Rouse, W.B. 1980. Computer-based manuals for procedural information. *IEEE Transactions on Systems, Man, and Cybernetics*, 10, pp. 506-510.

Rouse, W.B. 1987. Model-based evaluation of an integrated support system concept. *Large Scale Systems*, 13, pp. 33-42.

Rouse, W.B. 1991. *Design for Success: A Human-centered Approach to Designing Successful Products and Systems.* New York: Wiley.

Rouse, W.B., Cody, W.J., Boff, K.R., and Frey, P.R. 1990. Information systems for supporting design of complex human-machine systems. In C.T. Leondes, (Ed.), *Control and Dynamic Systems: Advances in Industrial and Aeronautical Systems.* Orlando, FL: Academic Press.

Rouse, W.B. and Hammer, J.M. 1990. *Assessing the Impact of Modeling Limits on Intelligent Systems.* Hampton, VA: NASA Langley Research Center.

Rouse, W.B. and Morris, N.M. 1986. On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100, pp. 349-363.

Schwarz, E., Beldie, I.P., and Pastoor, S. 1983. A comparison of paging and scrolling for changing screen contents by inexperienced users. *Human Factors*, Vol. 25, pp. 279-282.

Sewell, D.R., Rouse, W.B., and Johnson, W.B. 1989. *Initial Evaluation of Principles for Graphical Displays in Maintenance Problem Solving.* Arlington, VA: Office of Naval Research.

Smillie, R.J., Nugent, W.A., Sander, S.I., and Johnson, D.M. 1988. *A Comparitive Assessment of Paper-based and Computer-based Maintenance Information Delivery Systems* (NPRDC TR88-14). San Diego, CA: Navy Personnel Research and Development Center.

Smith, S.L. and Mosier, J.N. 1986. *Guidelines for Designing User Interface Software* (Tech. Rept. 86-278). Hanscom AFB, MA: Air Force Systems Command, Electronic System Division.

Tullis, T.S. 1986. Display Analysis Program: Version 4.0 (User's Guide). Unpublished program and user's guide.

Tullis, T.S. 1983. The formatting of alphanumeric displays: A review and analysis. *Human Factors*, Vol. 25, pp. 657-682.

APPENDIX A

ANOVA Results for Experiments One Through Five

MEASURE	SOURCE	SS	DF	MS	F	P
NUMBER OF DISPLAYS	Material	6.585	2	3.292	18.491	0.000
	Experience x Material	1.899	2	0.950	5.334	0.022
	Error	2.137	12	0.178		

Figure A-1. ANOVA Results for Experiment One (n = 6).

MEASURE	SOURCE	SS	DF	MS	F	P
USE OF DISPLAY TYPE						
High Abstraction	Within Subjects Effects:					
(Flow Diagrams)	Task Type	2.203	2	1.102	22.302	0.000
	Error	0.790	16	0.049		
Medium Abstraction	Within Subjects Effects:					
(Schematic Diagrams)	Task Type	1.110	2	0.555	8.723	0.003
	Error	1.017	16	0.064		
Low Abstraction	Within Subjects Effects:					
(Location Diagrams)	Task Type	5.669	2	2.834	26.763	0.000
	Error	1.695	16	0.106		
High Aggregation	Within Subjects Effects:					
	Task Type	3.270	2	1.635	32.610	0.000
	Error	0.802	16	0.050		
Medium Aggregation	Within Subjects Effects:					
	Task Type	1.428	2	0.714	7.644	0.005
	Error	1.494	16	0.093		
Low Aggregation	Within Subjects Effects:					
	Task Type	6.421	2	3.210	25.223	0.000
	Error	2.036	16	0.127		
Inappropriate Displays	Within Subjects Effects:					
	Task Type	0.076	2	0.038	5.727	0.013
	Error	0.106	16	0.007		
TOTAL TIME	Between Subjects Effects:					
	Experience	412200.223	1	412200.223	15.862	0.004
	Error	207891.676	8	25986.460		

Figure A-2. ANOVA Results for Experiment Two (n = 10).

MEASURE	SOURCE	SS	DF	MS	F	P
USE OF DISPLAY TYPE						
High Abstraction (Flow Diagrams)	Within Subjects Effects:					
	Task Type	10.121	2	5.061	29.391	0.000
	Error	3.788	22	0.172		
	2-Way Interactions:					
	Task Type x Experience	1.733	2	0.867	5.033	0.016
	Error	3.788	22	0.172		
Medium Abstraction (Schematic Diagrams)	Within Subjects Effects:					
	Task Type	5.041	2	2.520	10.012	0.001
	Error	5.538	22	0.252		
	2-Way Interactions:					
	Task Type x Experience	1.381	2	0.691	2.744	0.086
	Error	5.538	22	0.252		
Low Abstraction (Location Diagrams)	Within Subjects Effects:					
	Task Type	4.262	2	2.131	15.041	0.000
	Error	3.117	22	0.142		
High Aggregation	Within Subjects Effects:					
	Task Type	2.054	2	1.027	9.820	0.001
	Error	2.301	22	0.105		
Medium Aggregation	Within Subjects Effects:					
	Task Type	7.544	2	3.772	13.923	0.000
	Error	5.960	22	0.271		
Low Aggregation	Within Subjects Effects:					
	Task Type	2.808	2	1.404	10.935	0.001
	Error	2.824	22	0.128		
Inappropriate Displays	Within Subjects Effects:					
	Task Type	0.012	2	0.006	3.365	0.053
	Error	0.041	22	0.002		

Figure A-3. ANOVA Results for Experiment Three (n = 13).

MEASURE	SOURCE	SS	DF	MS	F	P
TOTAL TIME	Within Subjects Effects:					
	Task Type	284113.437	2	142056.719	23.314	0.000
	Error	134049.800	22	6093.173		

Figure A-3. ANOVA Results for Experiment Three (n = 13) (cont.).

MEASURE	SOURCE	SS	DF	MS	F	P
MAINTENANCE PERFORMANCE	2-Way Interactions:					
	Experience x Training	2.044	1	2.044	8.781	0.012
	Error	2.793	12	0.233		
	Experience x Flows Available	1.349	1	1.349	4.530	0.055
	Error	3.573	12	0.298		
	Low Experienced Subjects:					
	Between Subjects Effects:					
	Training	2.050	1	2.050	9.018	0.024
	Error	1.364	6	0.227		
	Within Subjects Effects:					
	Flows Available	1.484	1	1.484	5.749	0.053
	Error	1.549	6	0.258		
TIME TO SOLUTION	Between Subjects Effects:					
	Experience	7.972	1	7.972	11.819	0.005
	Error	8.094	12	0.674		
DIAGNOSTIC ERRORS	3-Way Interaction:					
	Experience x Training x Flows Available	0.251	1	0.251	4.940	0.046
	Error	0.609	12	0.051		

Figure A-4. ANOVA Results for Experiment Four (n = 16).

MEASURE	SOURCE	SS	DF	MS	F	P
USE OF DISPLAY TYPE						
Medium Abstraction (Schematic Diagrams)	Between Subjects Effects:					
	Experience	0.071	1	0.071	7.239	0.020
	Error	0.118	12	0.010		
	Training	0.085	1	0.085	8.638	0.012
	Error	0.118	12	0.010		
	Within Subjects Effect:					
	Flows Available	0.167	1	0.167	19.318	0.001
	Error	0.104	12	0.009		
Low Abstraction (Location Diagrams)	Between Subjects Effects:					
	Experience	0.019	1	0.019	17.093	0.001
	Error	0.013	12	0.001		
	Training	0.022	1	0.022	20.389	0.001
	Error	0.013	12	0.001		
NUMBER OF DISPLAYS ACCESSED	Between Subjects Effects:					
	Experience	4.459	1	4.459	14.515	0.002
	Error	3.686	12	0.307		
	Training	4.123	1	4.123	13.423	0.003
	Error	3.686	12	0.307		

Figure A-4. ANOVA Results for Experiment Four (n = 16) (cont.).

MEASURE	SOURCE	SS	DF	MS	F	P
MAINTENANCE PERFORMANCE	Within Subjects Effects:					
	Display Condition	5.948	2	2.974	3.847	0.041
	Error	13.914	18	0.773		
TIME TO SOLUTION	Within Subjects Effects:					
	Display Condition	7.783	2	3.892	7.462	0.004
	Error	9.387	18	0.522		
NUMBER OF DISPLAYS	Within Subjects Effects:					
ACCESSED	Display Condition	10.237	2	5.118	12.408	0.000
	Error	7.425	18	0.413		
USE OF DISPLAY TYPE						
High Abstraction	Within Subjects Effects:	0.267	1	0.267	19.440	0.002
(Flow Diagrams)	Display Condition	0.124	9	0.014		
	Error					
Medium Abstraction	Within Subjects Effects:					
(Schematic Diagrams)	Display Condition	0.405	2	0.203	15.172	0.000
	Error	0.241	18	0.013		
Medium Aggregation	Within Subjects Effects:					
	Display Condition	2.756	2	1.378	41.839	0.000
	Error	0.593	18	0.033		
Low Aggregation	Within Subjects Effects:					
	Display Condition	2.670	2	1.335	33.832	0.000
	Error	0.710	18	0.039		

Figure A-5. ANOVA Results for Experiment Five (n = 10).

Technical Report Distribution List - Manpower, Personnel, and Training R&D Program

As of 3/30/92
One copy to each addressee except as noted:

Director, Contract Research Department
Office of Naval Research (Code 11)
Arlington, VA 22217-5000

Chairman, MPT R&D Committee
Office of the Chief of Naval Research
Code 222
Arlington, VA 22217-5000

Program Manager, Operations
Research (Code 1111MA)
Office of Naval Research
Arlington, VA 22217-5000

Director, Life Sciences (Code 114)
Office of Naval Research
Arlington, VA 22217-5000

Director, Cognitive & Neural Sciences
(Code 1142)
Office of Naval Research
Arlington, VA 22217-5000

Cognitive Science (Code 1142CS)
Office of Naval Research
Arlington, VA 22217-5000

Perceptual Science (Code 1142PS)
Office of Naval Research
Arlington, VA 22217-5000

Defense Technical Information Center*
DTIC/DDA-2
Cameron Station, Building 5
Alexandria, VA 22314

CDR J. S. Hanna, Office of the Deputy
Asst. Secretary of the Navy (Manpower)
5D800, The Pentagon
Washington, DC 20350-1000

Head, Manpower, Personnel, and
Training Branch
Office of the CNO (Op-813)
4A478, The Pentagon
Washington, DC 20350-1000

Assistant for Manpower and Training
Office of the CNO (Op-911H)
5D772, The Pentagon
Washington, DC 20350-2000

Assistant for Planning and Technology
Development
Office of the DCNO(MPT) (Op-01B2)
Department of the Navy, AA-1822
Washington, DC 20350-2000

Deputy Director Total Force Training
and Education Division
Office of the DCNO(MPT) (Op-11B)
Department of the Navy
Washington, DC 20370-2000

R&D Coordinator, Attn: Jan Hart
Office of the DCNO(MPT) (Op-11K1)
Department of the Navy, AA-G817
Washington, DC 20370-2000

Deputy Director Military Personnel
Policy Division
Office of the DCNO(MPT) (Op-13B)
Department of the Navy, AA-1825
Washington, DC 20370-2000

Head, Military Compensation Policy
Branch
Office of the DCNO(MPT) (Op-134)
Department of the Navy, AA-2837
Washington, DC 20370-2000

*Note: 2 copies go to DTIC

Headquarters U.S. Marine Corps
Code MA
Washington, DC 20380-0001

Head, Leadership Branch
Naval Military Personnel Command
Attn: LT Gary Kent, NMPC-621
Department of the Navy, AA-1603
Washington, DC 20370-5620

Director, Research & Analysis Division
Navy Recruiting Command (Code 223)
4015 Wilson Boulevard, Room 215
Arlington, VA 22203-1991

Technical Director, Dr. John Silva
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

Dr. Kobus
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

Head, Human Factors Division
Naval Training Systems Ctr. (Code 26)
12350 Research Parkway
Orlando, FL 32826-3224

Naval Training Systems Center
ATTN: Dr. Eduardo Salas, (Code 262)
12350 Research Parkway
Orlando, FL 32826-3224

Naval Training Systems Center
ATTN: Dr. Robert Hays, (Code 262)
12350 Research Parkway
Orlando, FL 32826-3224

Naval Training Systems Center
ATTN: Dr. Rosemary Garris, (262)
12350 Research Parkway
Orlando, FL 32826-3224

Technical Director
Navy Personnel R&D Center
Code 01
San Diego, CA 92152-6800

Director, Manpower Systems Dept.
Code 11
Navy Personnel R&D Center
San Diego, CA 92152-6800

Director, Personnel Systems Dept.
Code 12
Navy Personnel R&D Center
San Diego, CA 92152-6800

Director, Testing Systems Department
Navy Personnel R&D Center
Code 13
San Diego, CA 92152-6800

Director, Training Systems Dept.
Code 14
Navy Personnel R&D Center
San Diego, CA 92152-6800

Director, Training Technology Dept.
Code 15
Navy Personnel R&D Center
San Diego, CA 92152-6800

Director, Organizational Systems Dept.
Code 16
Navy Personnel R&D Center
San Diego, CA 92152-6800

Naval Ocean Systems Center
Command Support Technology Division
Attn: Mr. Jeffrey Grossman, Code 44
San Diego, CA 92152-5000

Chairman, Dept. of Admin. Sciences
(Code AS)
Naval Postgraduate School
Monterey, CA 93943-5100

Chairman, Department of Operations
Research (Code OR)
Naval Postgraduate School
Monterey, CA 93943-5100

Director, Instructional Development and
Educational Program Support Dept.
Naval Education and Training Program
Management Support Activity
(NETPMSA)
Pensacola, FL 32509-5100

Academic Programs and Research
Branch
Naval Technical Training Command
Code N6, NAS Memphis, Bldg. C-1
Millington, TN 38054-5056

Director, Defense Personnel Security
Research and Education Center
Suite E, Building 455
99 Pacific Street
Monterey, CA 93940-2481

Technical Director
U.S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Chief Scientist for Human Resources
The Armstrong Laboratory -- AL/HR - CA
Brooks AFB, TX 78235 - 5601

Director, Manpower & Training Program
Center for Naval Analyses
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Library- Code 231
Navy Personnel R&D Center
San Diego, CA 92152-6800

Library
Naval War College
Newport, RI 02940

Chief, Survey and Market
Analysis Division
Defense Manpower Data Center
1600 Wilson Boulevard, #400
Arlington, VA 22209-2593

Program Director
Manpower Research & Advisory Services
Smithsonian Institution
801 North Pitt Street, Suite 120
Alexandria, VA 22314-1713

Scientific Management Associates, Inc.
Theodore Post
Senior Professional
8201 Corporate Drive
Suite 800
Landover, MD 20785

FAA Technical Center
Dr. Ed Buckley
Code ACD 305
Atlantic City Airport
Atlantic City, NJ 08405

Dr. Stanley C. Collyer
Technology Area Manager
Manpower, Personnel, and Training Programs
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Robert J. Smillie
Human Factors Research
Naval Ocean Systems Center
Code 443
San Diego, CA 92152-5000